Notes on Matrices and Matrix Operations
Isaiah Lankham, Bruno Nachtergaele, Anne Schilling

(February 4, 2007)

1   Definition of and Notation for Matrices

Let $m, n \in \mathbb{Z}_+$ be positive integers. Then we begin by defining an $m \times n$ *matrix* A to be a rectangular array of numbers

$$A = (a_{ij})_{i,j=1}^{m,n} = (A^{(i,j)})_{i,j=1}^{m,n} = \left. \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \right\} m \ \text{ numbers}$$

$$\underbrace{}_{n \ \text{ numbers}}$$

where each element $a_{ij} \in \mathbb{F}$ in the array is called an *entry* of A (specifically, a_{ij} is called the "i, j entry"), i indexes the *rows* of A by ranging over the set $\{1, \ldots, m\}$, and j indexes the *columns* of A by ranging over the set $\{1, \ldots, n\}$. We say that the matrix A has *size* $m \times n$ and note that it is a (finite) sequence of doubly-subscripted numbers for which the two subscripts in no way depend upon each other.

Given the ubiquity of matrices in mathematics thought, a rich vocabulary has been developed for describing various properties and features of matrices that are most useful to their application. In addition, there is also a rich set of equivalent notations. For the purposes of these notes, we will use the above notation unless the size of the matrix is understood from context or is unimportant. In this case, we will drop much of this notation and denote a matrix simply as

$$A = (a_{ij}) \quad \text{or} \quad A = (a_{ij})_{m \times n}.$$

To get a sense of the essential vocabulary, suppose that we have an $m \times n$ matrix $A = (a_{ij})$ with $m = n$. Then we call A a *square* matrix. The elements $a_{11}, a_{22}, \ldots, a_{nn}$ in a square matrix form what is called the *main diagonal* of A, and the elements $a_{1n}, a_{2,n-1}, \ldots, a_{n1}$ form what is sometimes called the *skew main diagonal* of A. Entries not on the main diagonal are also often called *off-diagonal* entries, and a matrix whose off-diagonal entries are all zero is called a *diagonal matrix*. It is common to call the elements $a_{12}, a_{23}, \ldots, a_{n-1,n}$ the *superdiagonal* of A and $a_{21}, a_{32}, \ldots, a_{n,n-1}$ the *subdiagonal*. The motivation for this

terminology should be clear if you create a sample square matrix and trace the entries within these particular subsequences of the matrix.

Square matrices are important because they are fundamental to applications of Linear Algebra. In particular, virtually every use of Linear Algebra in problem solving either involves square matrices directly or employ them in some indirect manner. In addition, virtually every usage also involves the notion of *vector*, where here we mean either an $m \times 1$ matrix (a.k.a. a *row vector*) or a $1 \times n$ matrix (a.k.a. a *column vector*).

Example 1.1. Suppose that $A = (a_{ij})$, $B = (b_{ij})$, $C = (c_{ij})$, $D = (d_{ij})$, and $E = (e_{ij})$ are the following matrices over $\mathbb{F}$:

$$A = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}, \; B = \begin{bmatrix} 4 & -1 \\ 0 & 2 \end{bmatrix}, \; C = \begin{bmatrix} 1, & 4, & 2 \end{bmatrix}, \; D = \begin{bmatrix} 1 & 5 & 2 \\ -1 & 0 & 1 \\ 3 & 2 & 4 \end{bmatrix}, \; E = \begin{bmatrix} 6 & 1 & 3 \\ -1 & 1 & 2 \\ 4 & 1 & 3 \end{bmatrix}.$$

Then we say that A is a 3×1 matrix (a.k.a a column vector), B is a 2×2 square matrix, C is a 1×3 matrix (a.k.a. a row vector), and both D and E are square 3×3 matrices.

We can discuss individual entries in each matrix. E.g., $d_{12} = 5$ and $e_{12} = e_{22} = e_{32} = 1$. The diagonal of D is the sequence $d_{11} = 1, d_{22} = 0, d_{33} = 4$. The subdiagonal of E is the sequence $e_{21} = -1, e_{32} = 1$.

We also note that B is called an *upper-triangular* matrix since all of the elements "below" the main diagonal are zero. However, none of the matrices above are diagonal matrices. Given any positive integer $n \in \mathbb{Z}_+$, we can construct the diagonal matrices I_n (called the *identity matrix*) and $0_{n \times n}$ (called the *zero matrix*) by setting

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \quad \text{and} \quad 0_{n \times n} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix},$$

where each of these matrices is understood to be a square matrix of size $n \times n$. The zero matrix $0_{m \times n}$ is analogously defined for any two positive integer $m, n \in \mathbb{Z}_+$ and has size $m \times n$.

2 Matrix Arithmetic

Given positive integers $m, n \in \mathbb{Z}_+$, we use $\mathbb{F}^{m \times n}$ to denote the set of all $m \times n$ matrices having entries over $\mathbb{F}$. In this section, we examine algebraic properties of this set. Specifically, $\mathbb{F}^{m \times n}$ forms a vector space under the operations of component-wise addition and scalar multiplication, and it is isomorphic to $\mathbb{F}^{mn}$ as a vector space.

We also define a multiplication operation between matrices of compatible size and show that this multiplication operation interacts with the vector space structure on $\mathbb{F}^{m \times n}$ in a natural way.

2.1 Addition and Scalar Multiplication

Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ matrices over $\mathbb{F}$ (where $m, n \in \mathbb{Z}_+$), and let $\alpha \in \mathbb{F}$. Then *matrix addition* $A + B = ((a + b)_{ij})_{m \times n}$ and *scalar multiplication* $\alpha A = ((\alpha a)_{ij})_{m \times n}$ are both defined component-wise, meaning

$$(a + b)_{ij} = a_{ij} + b_{ij} \quad \text{and} \quad (\alpha a)_{ij} = \alpha a_{ij}.$$

Equivalently, $A + B$ is the $m \times n$ matrix given by

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix},$$

and αA is the $m \times n$ matrix given by

$$\alpha \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{bmatrix}.$$

Example 2.1. With notation as in Example 1.1 above,

$$D + E = \begin{bmatrix} 7 & 6 & 5 \\ -2 & 1 & 3 \\ 7 & 3 & 7 \end{bmatrix},$$

and no two other matrices from Example 1.1 can be added since their sizes are not compatible. We can similar make calculations like

$$D - E = D + (-1 \cdot E) = \begin{bmatrix} -5 & 4 & -1 \\ 0 & -1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad 0D = 0E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0_{3 \times 3}.$$

It is important to note that while these are not the only ways of defining addition and scalar multiplication operations on $\mathbb{F}^{m \times n}$, the above operations have the advantage of endowing $\mathbb{F}^{m \times n}$ with a reasonably natural vector space structure. As a vector space, $\mathbb{F}^{m \times n}$ is seen to have dimension mn since we can build the *standard basis matrices*

$$E_{11}, E_{12}, \ldots, E_{1n}, E_{21}, E_{22}, \ldots, E_{2n}, \ldots, E_{m1}, E_{m2}, \ldots, E_{mn}$$

by analogy to the standard basis for $\mathbb{F}^{mn}$. That is, each $E_{k\ell} = ((e^{(k,\ell)})_{ij})$ satisfies

$$(e^{(k,\ell)})_{ij} = \begin{cases} 1, & \text{if } i = k \text{ and } j = \ell \\ 0, & \text{otherwise} \end{cases}.$$

This allows us to build a vector space isomorphism $\mathbb{F}^{m \times n} \to \mathbb{F}^{mn}$ using a bijection that simply "lays each matrix out flat". In other words, given $A = (a_{ij}) \in \mathbb{F}^{m \times n}$,

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \mapsto (a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, a_{22}, \ldots, a_{2n}, \ldots, a_{m1}, a_{m2}, \ldots, a_{mn}) \in \mathbb{F}^{mn}.$$

Example 2.2. The vector space $\mathbb{R}^{2 \times 3}$ of 2×3 matrices over $\mathbb{R}$ has standard basis

$$\left\{ E_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, E_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \right.$$

$$\left. E_{21} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, E_{22} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, E_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\},$$

which is seen to naturally correspond with the standard basis $\{e_1, \ldots, e_6\}$ for $\mathbb{R}_6$, where

$$e_1 = (1, 0, 0, 0, 0, 0), e_2 = (0, 1, 0, 0, 0, 0), \ldots, e_6 = (0, 0, 0, 0, 0, 1).$$

Of course, it is not enough to just assert that $\mathbb{F}^{m \times n}$ is a vector space since we have yet to verify that the above defined operations of addition and scalar multiplication satisfy the axioms of a vector space. The proof of the following theorem is straight-forward and something that you should work through for practice with matrix notation.

Theorem 2.3. *Given positive integers $m, n \in \mathbb{Z}_+$ and the operations of matrix addition and scalar multiplication as defined above, the set $\mathbb{F}^{m \times n}$ of all $m \times n$ matrices satisfies each of the following properties.*

1. *(associativity of matrix addition) Given any three matrices $A, B, C \in \mathbb{F}^{m \times n}$,*

$$(A + B) + C = A + (B + C).$$

2. *(additive identity for matrix addition) Given any matrices $A \in \mathbb{F}^{m \times n}$,*

$$A + 0_{m \times n} = 0_{m \times n} + A = A.$$

3. *(additive inverses for matrix addition) Given any matrices $A \in \mathbb{F}^{m \times n}$, there exists a matrix $-A \in \mathbb{F}^{m \times n}$ such that*

$$A + (-A) = (-A) + A = 0_{m \times n}.$$

4

4. *(commutativity of matrix addition) Given any two matrices $A, B \in \mathbb{F}^{m \times n}$,*

$$A + B = B + A.$$

5. *(associativity of scalar multiplication) Given any matrix $A \in \mathbb{F}^{m \times n}$ and any two scalars $\alpha, \beta \in \mathbb{F}$,*

$$(\alpha \beta)A = \alpha(\beta A).$$

6. *(multiplicative identity for scalar multiplication) Given any matrix $A \in \mathbb{F}^{m \times n}$ and denoting by 1 the multiplicative identity of $\mathbb{F}$,*

$$1A = A.$$

7. *(distributivity of scalar multiplication) Given any two matrices $A, B \in \mathbb{F}^{m \times n}$ and any two scalars $\alpha, \beta \in \mathbb{F}$,*

$$(\alpha + \beta)A = \alpha A + \beta A \quad and \quad \alpha(A + B) = \alpha A + \alpha B.$$

As a consequence of Theorem 2.3, every property that holds for an arbitrary vector space can be taken as a property of $\mathbb{F}^{m \times n}$ specifically. Some of the more basic such properties are collected as follows.

Corollary 2.4. *Given positive integers $m, n \in \mathbb{Z}_+$ and the operations of matrix addition and scalar multiplication as defined above, the set $\mathbb{F}^{m \times n}$ of all $m \times n$ matrices satisfies each of the following properties.*

1. *Given any matrix $A \in \mathbb{F}^{m \times n}$, given any scalar $\alpha \in \mathbb{F}$, and denoting by 0 the additive identity of $\mathbb{F}$,*

$$0A = A \quad and \quad \alpha 0_{m \times n} = 0_{m \times n}.$$

2. *Given any matrix $A \in \mathbb{F}^{m \times n}$ and any scalar $\alpha \in \mathbb{F}$,*

$$\alpha A = 0 \implies either \quad \alpha = 0 \quad or \quad A = 0_{m \times n}.$$

3. *Given any matrix $A \in \mathbb{F}^{m \times n}$ and any scalar $\alpha \in \mathbb{F}$,*

$$-(\alpha A) = (-\alpha)A = \alpha(-A).$$

In particular, the additive inverse $-A$ of A is given by $-A = (-1)A$, where 1 denoted the additivity identity for $\mathbb{F}$.

While one could prove Corollary 2.4 directly from definitions, the point of recognizing $\mathbb{F}^{m \times n}$ as a vector space is that you get to use these results without worrying about their proof. Moreover, there is no need to separately prove that they hold for both $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$.

2.2 Multiplying Matrices

Let $r, s, t \in \mathbb{Z}_+$ be positive integers, $A = (a_{ij}) \in \mathbb{F}^{r \times s}$ be an $r \times s$ matrix, and $B = (b_{ij}) \in \mathbb{F}^{s \times t}$ be an $s \times t$ matrix. Then *matrix multiplication* $AB = ((ab)_{ij})_{r \times t}$ is defined by

$$(ab)_{ij} = \sum_{k=1}^{s} a_{ik} b_{kj}.$$

In particular, note that the "i, j entry" of the matrix product AB involves a summation in the positive integer s, where s is both the number of columns in A and the number of rows in B. Thus, this multiplication is only defined for matrices having equal "middle" dimension:

$$(a_{ij})_{r \times s}(b_{ij})_{s \times t} = r \left\{ \underbrace{\begin{bmatrix} a_{11} & \cdots & a_{1s} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rs} \end{bmatrix}}_{s} \underbrace{\begin{bmatrix} b_{11} & \cdots & b_{1t} \\ \vdots & \ddots & \vdots \\ b_{s1} & \cdots & b_{st} \end{bmatrix}}_{t} \right\} s = \underbrace{\begin{bmatrix} \sum_{k=1}^{s} a_{1k} b_{k1} & \cdots & \sum_{k=1}^{s} a_{1k} b_{kt} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{s} a_{rk} b_{k1} & \cdots & \sum_{k=1}^{s} a_{rk} b_{kt} \end{bmatrix}}_{t} \Bigg\} r$$

Alternatively, if we let $n \in \mathbb{Z}_+$ be a positive integer, then another way of viewing matrix multiplication is through the use of the standard inner product on $\mathbb{F}^n = \mathbb{F}^{1 \times n} = \mathbb{F}^{n \times 1}$. In particular, we define the *dot product* (a.k.a. *Euclidean inner product*) of the row vector $x = (x_{1j}) \in \mathbb{F}^{1 \times n}$ and the column vector $y = (y_{i1}) \in \mathbb{F}^{n \times 1}$ to be

$$x \cdot y = \begin{bmatrix} x_{11}, & \cdots, & x_{1n} \end{bmatrix} \cdot \begin{bmatrix} y_{11} \\ \vdots \\ y_{n1} \end{bmatrix} = \sum_{k=1}^{n} x_{1k} y_{k1} \in \mathbb{F}.$$

Moreover, we can decompose matrices $A = (a_{ij})_{r \times s}$ and $B = (b_{ij})_{s \times t}$ into their constitute row vectors by fixing a positive integer $k \in \mathbb{Z}_+$ and setting

$$A^{(k, \cdot)} = \begin{bmatrix} a_{k1}, & \cdots, & a_{ks} \end{bmatrix} \in \mathbb{F}^{1 \times s} \quad \text{and} \quad B^{(k, \cdot)} = \begin{bmatrix} b_{k1}, & \cdots, & b_{kt} \end{bmatrix} \in \mathbb{F}^{1 \times t}.$$

Similarly, fixing $\ell \in \mathbb{Z}_+$, we can also decompose them into column vectors

$$A^{(\cdot, \ell)} = \begin{bmatrix} a_{1\ell} \\ \vdots \\ a_{r\ell} \end{bmatrix} \in \mathbb{F}^{r \times 1} \quad \text{and} \quad B^{(\cdot, \ell)} = \begin{bmatrix} b_{1\ell} \\ \vdots \\ b_{s\ell} \end{bmatrix} \in \mathbb{F}^{s \times 1}.$$

It follows that the product AB is then the follow matrix of dot products:

$$AB = \begin{bmatrix} A^{(1, \cdot)} \cdot B^{(\cdot, 1)} & \cdots & A^{(1, \cdot)} \cdot B^{(\cdot, t)} \\ \vdots & \ddots & \vdots \\ A^{(r, \cdot)} \cdot B^{(\cdot, 1)} & \cdots & A^{(r, \cdot)} \cdot B^{(\cdot, t)} \end{bmatrix} \in \mathbb{F}^{r \times t}.$$

Example 2.5. With notation as in Example 1.1 above, you should sit down and use the above definitions in order to verify that the following matrix products hold.

$$AC = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 1, & 4, & 2 \end{bmatrix} = \begin{bmatrix} 3 & 12 & 6 \\ -1 & -4 & -2 \\ 1 & 4 & 2 \end{bmatrix} \in \mathbb{F}^{3 \times 3},$$

$$CA = \begin{bmatrix} 1, & 4, & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = 3 - 4 + 2 = 1 \in \mathbb{F},$$

$$B^2 = BB = \begin{bmatrix} 4 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 16 & -6 \\ 0 & 4 \end{bmatrix} \in \mathbb{F}^{2 \times 2},$$

$$CE = \begin{bmatrix} 1, & 4, & 2 \end{bmatrix} \begin{bmatrix} 6 & 1 & 3 \\ -1 & 1 & 2 \\ 4 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 10, & 7, & 17 \end{bmatrix} \in \mathbb{F}^{1 \times 3}, \text{ and}$$

$$DA = \begin{bmatrix} 1 & 5 & 2 \\ -1 & 0 & 1 \\ 3 & 2 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 11 \end{bmatrix} \in \mathbb{F}^{3 \times 1}.$$

Note, though, that B cannot be multiplied by any of the other matrices, nor does it make sense to try to form the products AD, AE, DC, and EC due to the inherent size mismatches.

As illustrated in Example 2.5 above, matrix multiplication is not a commutative operation (since, e.g., $AC \in \mathbb{F}^{3 \times 3}$ while $CA \in \mathbb{F}^{1 \times 1}$). Nonetheless, despite the complexity of its definition, the matrix product otherwise satisfies many familiar properties of a multiplication operation. We summarize the most basic of these properties in the following theorem.

Theorem 2.6. *Let $r, s, t, u \in \mathbb{Z}_+$ be positive integers.*

1. *(associativity of matrix multiplication) Given $A \in \mathbb{F}^{r \times s}$, $B \in \mathbb{F}^{s \times t}$, and $C \in \mathbb{F}^{t \times u}$,*

$$A(BC) = (AB)C.$$

2. *(distributivity of matrix multiplication) Given $A \in \mathbb{F}^{r \times s}$, $B, C \in \mathbb{F}^{s \times t}$, and $D \in \mathbb{F}^{t \times u}$,*

$$A(B + C) = AB + AC \quad and \quad (B + C)D = BD + CD.$$

3. *(compatibility with scalar multiplication) Given $A \in \mathbb{F}^{r \times s}$, $B \in \mathbb{F}^{s \times t}$, and $\alpha \in \mathbb{F}$,*

$$\alpha(AB) = (\alpha A)B = A(\alpha B).$$

As with Theorem 2.3, you should work through a proof of each part of Theorem 2.6 (and especially of the first part) in order to practice manipulating the indices of entries correctly.

7

At the same time, you should be careful to not blithely perform operations on matrices as you would with numbers. The fact that matrix multiplication is not a commutative operation should make it clear that significantly more care is required with matrix arithmetic. As another example, given a positive integers $n \in \mathbb{Z}_+$, the set $\mathbb{F}^{n \times n}$ has what are called *zero divisors*. That is, there exist non-zero matrices $A, B \in \mathbb{F}^{n \times n}$ such that $AB = 0_{n \times n}$:

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0_{2 \times 2}.$$

Moreover, note that there exist matrices $A, B, C \in \mathbb{F}^{n \times n}$ such that $AB = AC$ but $B \neq C$:

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

As a result, we say that the set $\mathbb{F}^{n \times n}$ fails to have the so-called *cancellation property*. This failure is a direct result of the fact that there are non-zero matrices in $\mathbb{F}^{n \times n}$ that have no multiplicative inverse. We discuss matrix invertibility at length in the next section and define a special subset $GL(n, \mathbb{F}) \subset \mathbb{F}^{n \times n}$ upon which the cancellation property does hold.

3   Some Special Operations on Square Matrices

Given a positive integer $n \in \mathbb{Z}_+$, the set of square matrices $\mathbb{F}^{n \times n}$ admits a natural multiplication operation between any two elements as defined in the section above, and so we can also discuss invertibility of elements in $\mathbb{F}^{n \times n}$ with respect to this multiplication operation. If you have taken a more introductory course in Linear Algebra, then you have most likely seen much of this material in one form or another. Criteria for determining when a matrix can be inverted and uses of invertibility typically dominate the bulk of such courses.

We also define three important operations on square matrices called the transpose, conjugate transpose, and the trace. These will then be seen to interact with matrix multiplication and invertibility to form special classes of matrices that are extremely important to applications of Linear Algebra.

3.1   Invertibility of Square Matrices

Given a positive integer $n \in \mathbb{Z}_+$, we say that a square matrices $A \in \mathbb{F}^{n \times n}$ is *invertible* (a.k.a. *nonsingular*) if there exists a square matrix $B \in \mathbb{F}^{n \times n}$ such that

$$AB = BA = I_n.$$

Moreover, we denote by $GL(n, \mathbb{F})$ the set of all invertible $n \times n$ matrices over $\mathbb{F}$. Note, in particular, that the zero matrix $0_{n \times n} \notin GL(n, \mathbb{F})$ so that $GL(n, \mathbb{F})$ *is not* a subspace of $\mathbb{F}^{n \times n}$.

One can prove that if the multiplicative inverse of a matrix A exists, then it is unique, and so we usually denote it by A^{-1}. This notation for matrix inverse is by analogy to the notation

for the multiplicative inverse of a scalar and should not be taken to imply that it is possible to "divide" by a matrix. Since matrix multiplication is not a commutative operation, care must be taken to check whether or not algebraic properties for scalars continue to hold. We summarize the most basic of these properties as follows.

Theorem 3.1. *Let $n \in \mathbb{Z}_+$ be positive integers. Then, given any matrices $A, B \in GL(n, \mathbb{F})$,*

1. *the inverse matrix $A^{-1} \in GL(n, \mathbb{F})$ and satisfies $(A^{-1})^{-1} = A$.*

2. *the matrix power $A^m \in GL(n, \mathbb{F})$ and satisfies $(A^m)^{-1} = (A^{-1})^m$, where $m \in \mathbb{N}$ is any non-negative integer.*

3. *the matrix $\alpha A \in GL(n, \mathbb{F})$ and satisfies $(\alpha A)^{-1} = \alpha^{-1} A^{-1}$, where $0 \neq \alpha \in \mathbb{F}$ is any non-zero scalar.*

4. *the product $AB \in GL(n, \mathbb{F})$ and has inverse given by the formula*

$$(AB)^{-1} = B^{-1} A^{-1}.$$

Moreover, $GL(n, \mathbb{F})$ has the cancellation property. *In other words, given $A, B, C \in GL(n, \mathbb{F})$, if $AB = AC$, then $B = C$.*

At the same time, it is important to note that not all matrices are invertible. As an illustration of the subtlety involved in understanding invertibility, we give the following theorem for the 2×2 case.

Theorem 3.2. *Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in \mathbb{F}^{2 \times 2}$. Then A is invertible if and only if the expression $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Moreover, if A is invertible, then*

$$A^{-1} = \begin{bmatrix} \dfrac{a_{22}}{a_{11}a_{22} - a_{12}a_{21}} & \dfrac{-a_{12}}{a_{11}a_{22} - a_{12}a_{21}} \\ \dfrac{-a_{21}}{a_{11}a_{22} - a_{12}a_{21}} & \dfrac{a_{11}}{a_{11}a_{22} - a_{12}a_{21}} \end{bmatrix}.$$

A more general theorem holds for larger matrices, but its statement requires machinery (namely the notion of the determinant of a matrix) that is outside the scope of these notes.

We close this section by noting that the set $GL(n, \mathbb{F})$ of all invertible $n \times n$ matrices over $\mathbb{F}$ is often called the *general linear group*. This set has so many important uses in mathematics that there are many equivalent notations for it, including $GL_n(\mathbb{F})$, $GL(\mathbb{F}^n)$, and simply $GL(n)$ or GL_n if it is not important to emphasis the dependence on $\mathbb{F}$. Moreover, the usage of the term "group" in the name "general linear group" is highly technical. To explain this term, we give the following definition, which is unequivocably one of the most fundamental and ubiquitous notions in all of mathematics.

Definition 3.3. Let G be a set and $*$ be a binary operation of G. (In other words, $*$ is the name of a function $* : G \times G \to G$, where the notation $a*b$ means the same thing as $*(a, b)$.) Then G is said to *form a group* under $*$ if the following three conditions are satisfied:

1. (associativity) Given any three elements $a, b, c \in G$, $(a * b) * c = a * (b * c)$.

2. (existence of an identity element) There is an element $e \in G$ such that, given any element $a \in G$, $a * e = e * a = a$.

3. (existence of inverse elements) Given any element $a \in G$, there is an element $b \in G$ such that $a * b = b * a = e$.

You should recognize these as three of the four axioms that addition in a vector space must satisfy. In particular, a vector space is said to be an *abelian group* under the binary operation of vector addition, where *abelian* is defined as follows.

Definition 3.4. Let G be a group under binary operation $*$. Then G is called an *abelian* (a.k.a. *commutative*) group if, given any two elements $a, b \in G$, $a * b = b * a$.

Note in particular that $GL(n, \mathbb{F}) \subset \mathbb{F}^{n \times n}$ and that $GL(n, \mathbb{F})$ is a non-abelian group under matrix multiplication when $n \geq 2$. At the same time, $\mathbb{F}^{n \times n}$ is an abelian group under vector addition (since it is a vector space as shown in Theorem 2.3) but does not form a group under matrix multiplication. The set $GL(n, \mathbb{F})$, similarly, does not form a group under vector addition since, e.g., the additive identity element $0_{n \times n} \notin GL(n, \mathbb{F})$. This illustrates the importance of emphasizing the operation under which we which to consider whether or not a set forms a mathematical structure such as a group.

The notion of a group is fundamental to mathematical thought because it is often used as a building block for more complicated algebraic structures, as in a vector space being an abelian group together with a "compatibly defined" scalar multiplication. As another example, we note that the set of real numbers $\mathbb{R}$ is an abelian group under addition. Moreover, the set of non-zero real numbers $\mathbb{R} \setminus \{0\}$ can also be seen to form a group under multiplication. Thus, since $\mathbb{R}$ is an abelian group under these two "compatibly defined" notions of addition and multiplication, we call $\mathbb{R}$ a *field*. The set of complex numbers $\mathbb{C}$ is also seen to form a field, and there are many other important examples of fields including the set $\mathbb{Q}$ of rational numbers. Note, though, that a set like $\mathbb{Z}$ is not a field since $\mathbb{Z} \setminus \{0\}$ does not form a group under multiplication. More precisely, $\mathbb{Z}$ fails to completely satisfy the following definition.

Definition 3.5. Let G be a set with binary operations $+$ and $*$ defined on G. Then we say that G *forms a field* under $+$ and $*$ if the following three conditions are satisfied:

1. The set G forms an abelian group under $+$.

2. The set $G \setminus \{0\}$ forms an abelian group under $*$, where 0 denotes the identity element of G as a group under $+$.

3. Given any three elements $a, b, c \in G$, $a * (b + c) = a * b + a * c$.

3.2 Transpose and Conjugate Transpose

Given positive integers $m, n \in \mathbb{Z}_+$ and any matrix $A = (a_{ij}) \in \mathbb{F}^{m \times n}$, we define the *transpose* $A^T = ((a^T)_{ij}) \in \mathbb{F}^{n \times m}$ and the *conjugate transpose* $A^* = ((a^*)_{ij}) \in \mathbb{F}^{n \times m}$ by

$$(a^T)_{ij} = a_{ji} \quad \text{and} \quad (a^*)_{ij} = \overline{a_{ji}},$$

where $\overline{a_{ji}}$ denotes the complex conjugate of the scalar $a_{ji} \in \mathbb{F}$. In particular, if $A \in \mathbb{R}^{m \times n}$, then note that $A^T = A^*$.

Example 3.6. With notation as in Example 1.1 above,

$$A^T = \begin{bmatrix} 3 & -1 & 1 \end{bmatrix}, \; B^T = \begin{bmatrix} 4 & 0 \\ -1 & 2 \end{bmatrix}, \; C^T = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}, \; D^T = \begin{bmatrix} 1 & -1 & 3 \\ 5 & 0 & 2 \\ 2 & 1 & 4 \end{bmatrix}, \; E^T = \begin{bmatrix} 6 & -1 & 4 \\ 1 & 1 & 1 \\ 3 & 2 & 3 \end{bmatrix}.$$

One of the motivations for defining the operations of transpose and conjugate transpose is that they interact with the usual arithmetic operations on matrices in a natural manner. We summarizes the most fundamental of these interactions in the following theorem.

Theorem 3.7. *Given positive integers $m, n \in \mathbb{Z}_+$ and any matrices $A, B \in \mathbb{F}^{m \times n}$,*

1. *$(A^T)^T = A$ and $(A^*)^* = A$.*

2. *$(A + B)^T = A^T + B^T$ and $(A + B)^* = A^* + B^*$.*

3. *$(\alpha A)^T = \alpha A^T$ and $(\alpha A)^* = \overline{\alpha} A^*$, where $\alpha \in \mathbb{F}$ is any scalar.*

4. *$(AB)^T = B^T A^T$.*

5. *if $m = n$ and $A \in GL(n)$, then $A^T, A^* \in GL(n)$ with respective inverses given by*

$$(A^T)^{-1} = (A^{-1})^T \quad \text{and} \quad (A^*)^{-1} = (A^{-1})^*.$$

Another motivation for defining the transpose and conjugate transpose operations is that they allow us to define several very special classes of matrices.

Definition 3.8. Given a positive integer $n \in \mathbb{Z}_+$, we say that the square matrix $A \in \mathbb{F}^{n \times n}$

1. is *symmetric* if $A = A^T$.

2. is *Hermitian* if $A = A^*$.

3. is *orthogonal* if $A \in GL(n, \mathbb{R})$ and $A^{-1} = A^T$. Moreover, we define the *(real) orthogonal group* to be the set $O(n) = \{A \in GL(n, \mathbb{R}) \mid A^{-1} = A^T\}$.

4. is *unitary* if $A \in GL(n, \mathbb{C})$ and $A^{-1} = A^*$. Moreover, we define the *(complex) unitary group* to be the set $U(n) = \{A \in GL(n, \mathbb{C}) \mid A^{-1} = A^*\}$.

11

It turns out that a lot can be said about the eigen-information specific to each of these special classes of matrices. For example, real symmetric and complex Hermitian matrices always have real eigenvalues. Moreover, given any matrix $A \in \mathbb{R}^{m \times n}$, AA^T is a symmetric matrix with real, non-negative eigenvalues. Similarly, for $A \in \mathbb{C}^{m \times n}$, AA^* is Hermitian with real, non-negative eigenvalues.

3.3 Trace

We conclude these notes by defining one more fundamental operation on square matrices. Specifically, given a positive integer $n \in \mathbb{Z}_+$ and any matrix $A = (a_{ij}) \in \mathbb{F}^{n \times n}$, we define the *trace* of A to be the scalar

$$\text{trace}(A) = \sum_{k=1}^{n} a_{kk} \in \mathbb{F}.$$

Example 3.9. With notation as in Example 1.1 above,

$$\text{trace}(B) = 4 + 2 = 6, \ \text{trace}(D) = 1 + 0 + 4 = 5, \ \text{ and } \ \text{trace}(E) = 6 + 1 + 3 = 10.$$

Note, in particular, that the traces of A and C are not defined since these are not square matrices.

We summarize some of the most basic properties of the trace operation in the following theorem, including its connection to the transpose operations defined in the previous section.

Theorem 3.10. *Given a positive integer $m, n \in \mathbb{Z}_+$ and square matrices $A, B \in \mathbb{F}^{n \times n}$,*

1. *$trace(\alpha A) = \alpha \, trace(A)$, for any scalar $\alpha \in \mathbb{F}$.*

2. *$trace(A + B) = trace(A) + trace(B)$.*

3. *$trace(A^T) = trace(A)$ and $trace(A^*) = \overline{trace(A)}$.*

4. *$trace(AA^*) = \sum\limits_{k=1}^{n} \sum\limits_{\ell=1}^{n} |a_{k\ell}|^2$. In particular, $trace(AA^*) = 0$ if and only if $A = 0_{n \times n}$.*

5. *$trace(AB) = trace(BA)$. More generally, given matrices $A_1, \ldots, A_m \in \mathbb{F}^{n \times n}$, the trace operation has the so-called* cyclic property, *meaning that*

$$trace(A_1 \cdots A_m) = trace(A_2 \cdots A_m A_1) = \cdots = trace(A_m A_1 \cdots A_{m-1}).$$

Moreover, if we define a linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ by setting $T(v) = Av$ for each $v \in \mathbb{F}^n$ and if T has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then $trace(A) = \sum\limits_{k=1}^{n} \lambda_k$.

4   Matrices in Context

As discussed in the introductory document, *What Is Linear Algebra?*, Linear Algebra can be seen as a branch of mathematics aimed at solving systems of linear equations involving a finite number of unknowns. As we will see below, matrices constitute one of the most fundamental tools for solving such systems. In particular, any arbitrary number of equations in any number of unknowns — as long as both are finite — can be encoded as a single matrix equation, and it is through such an encoding that computers are used on literally a daily basis to solve countless real world problems.

   In order to encode a system of linear equations as a matrix equation, we will first exploit the definition of matrix multiplication. Then we will see that this encoding directly relates to the notion of linear map between vector spaces.

4.1   Encoding Linear Systems as a Matrix Equation

Let $m, n \in \mathbb{Z}_+$ be positive integers, and consider the system of m linear equations in n unknowns $x_1, \ldots, x_n$,

$$\sum_{k=1}^{n} a_{1,k} x_k = b_1$$

$$\vdots$$

$$\sum_{k=1}^{n} a_{m,k} x_k = b_m,$$

where each $a_{i,j}, b_i \in \mathbb{F}$. We use $A = (a_{ij})$ to denote the $m \times n$ *coefficient matrix* associated to this linear system and $x = (x_i)$ to denote the $n \times 1$ column vector composed of the unknowns $x_1, \ldots, x_n$. In other words,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Using this notion, it should be clear from the definition of matrix multiplication that we have the equality

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = b,$$

where we have used b to denote the $m \times 1$ column vector formed from the right-hand side of the linear system.

4.2 Using Matrices to Define Linear Maps

As this course progresses, you see that encoding a linear system as a matrix equation is more than just a notational trick. Perhaps most fundamentally, this matrix point of view provides unparalleled insight into when a solution to a linear system is the only possible solution. (In general, the more than can be said with absolute certainty when solving a problem, the better.) We illustrate this in the following example.

Example 4.1. Suppose that we have the system of linear equations

$$2x_1 + x_2 = 0$$
$$x_1 - x_2 = 1,$$

where x_1 and x_2 are unknown real numbers. To solve this system, you have most likely been taught to first solve for one of the unknowns in one of the equations and then to substitute the result into the other equation. Here, for example, one might solve to obtain

$$x_1 = 1 + x_2$$

from the second equation. Then, substituting this in place of x_1 in the first equation, one obtains

$$2(1 + x_2) + x_2 = 0.$$

From this, we find that $x_2 = -2/3$. Then, by further substitution,

$$x_1 = 1 + \left(-\frac{2}{3}\right) = \frac{1}{3}.$$

Finally, if we wish to verify that this is the only solution to the given linear system, then we might appeal to a graph. In other words, since each of the equations corresponds to a line in the Euclidean plane $\mathbb{R}^2$, we can see that the solution $(x_1, x_2) = (1/3, -2/3)$ corresponds exactly to the single point of intersection between these two lines:
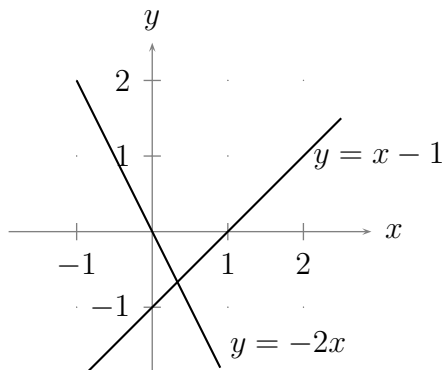
The above analysis, while tedious, is nonetheless straightforward. However, similar calculations can quickly become unwieldy when attempted on three or more equations, and

the graphical point of view is clearly impossible to generalize to dimensions four and higher. And yet, real world problems often require solutions to linear systems in thousands — if not millions or billions — of dimensions.

Fortunately, Linear Algebra provides a powerful and flexible tool for understanding linear systems, regardless of the dimension. This tool is the definition of linear map as is developed more carefully in the corresponding set of lecture notes. For now, you should just think of a linear map as a special type of function between vector spaces. The idea is that, by involving these special functions, we are able to work in an arbitrarily high number of dimensions with little more work than that of two dimensions. We illustrate this in the following example.

Example 4.2. Consider the matrix equation

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

which you should recognize as the matrix equation encoding of the linear system in Example 4.1. In other words, the column vector

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}$$

is equal to the column vector $b = \begin{bmatrix} 0, & 1 \end{bmatrix}^T$.

The idea is that the matrix equation above corresponds exactly to asking when $b \in \mathbb{R}^2$ is an element of the range of the function $L : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$L\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}.$$

Put another way, L is the function that, when given the column vector $x = \begin{bmatrix} x_1, & x_2 \end{bmatrix}^T$ as input, returns the column vector $\begin{bmatrix} 2x_1 + x_2, & x_1 - x_2 \end{bmatrix}^T$ as output.

Now, note that

$$L\left(\begin{bmatrix} 1/3 \\ -2/3 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and note also that L is a bijective function. Since L is bijective, this means that

$$x = \begin{bmatrix} 1/3 \\ -2/3 \end{bmatrix}$$

is the only possible input vector that can result in the output vector $\begin{bmatrix} 0, & 1 \end{bmatrix}^T$. Thus, we have verified the unique solution to the linear system of Example 4.1. More importantly, though, we have seen an example of a technique that trivially generalizes to verifying uniqueness of solutions to any number of equations.