Linear Algebra

Julian Dominic

04 November 2022

Preface

I wrote these notes principally FOR understanding, they are meant for future reference for a refresher on what I have learnt. As such, certain definitions may not be exactly precise but are rephrased into simpler terms.
I am currently using Schaum's Easy Outlines: Linear Algebra - Crash Course. ISBN 978-0-07-139880-0.
I plan to learn and use Linear Algebra by Hoffman and Kunze in the future. But I am currently clearing the pre-requisities for the book as the time of writing (04/11/2022); I am still learning proofs.

Contents

1 Vectors in $\mathbb{R}^n$

Although we will restrict ourselves in this chapter to vectors whose elements come from the field of real numbers, denoted by $\mathbb{R}$, many of our operations also apply to vectors whose entries come from arbitrary field $\boldsymbol{K}$. In the context of vectors, the elements of our number fields are called *scalars*.

1.1 List of Numbers

Suppose the height (in centimeters) of eight students are listed as follows:

$$155\ 165\ 175\ 185\ 180\ 170\ 160\ 150$$

One can denote all the values in the list using only one symbol, for example h, but with different subscripts; giving rise to h_i where $1 \leq i \leq 8$.

$$h_1\ h_2\ h_3\ h_4\ h_5\ h_6\ h_7\ h_8$$

It is easy to see that each subscript denotes the position of the value in the list; $h_1 = 155$, $h_2 = 165$, etc. Such a list of values, $\boldsymbol{h} = (h_1, h_2, h_3, \ldots, h_8)$ is called a *linear array* or *vector*.

The set of all n-tuples of real numbers, denoted by $\mathbb{R}^n$, is called *n-space*. A particular n-tuple in $\mathbb{R}^n$ like $\boldsymbol{u} = (a_1, a_2, \ldots, a_n)$ is called a *point* or *elements* of $\boldsymbol{u}$. Moreover, when discussing the space/domain of $\mathbb{R}^n$, we use the term *scalar* for the elements in $\mathbb{R}$ (notice that $n = 1$).

Consider two vectors, $\boldsymbol{u}$ and $\boldsymbol{v}$, are equal, we can express it as $\boldsymbol{u} = \boldsymbol{v}$ if they have the same number of elements and if the corresponding elements (position of the elements) are equal. Thi means that while vectors $(1, 2, 3)$ and $(2, 3, 1)$ have the same number of elements, the vectors are not equal because of the corresponding elements do not match each other from one vector to the other.

The vector $(0, 0, \underbrace{\ldots}_{n \text{ dots}}, 0)$; where all of the elements are zero, is called the *zero vector*, and it is usually denoted by $\mathbf{0}$.

1.2 Column Vectors

Sometimes a vector in n-space $\mathbb{R}^n$ is written vertically, rather than horizontally as shown previously. Such a vector is called a *column vector* while the ones shown previously are called *row vectors*. For example, the following are column vectors that belong to $\mathbb{R}^2$, $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^3$, respectively:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix} \quad \begin{bmatrix} 1.5 \\ \frac{2}{3} \\ -15 \end{bmatrix}$$

> **Remark 1.1.** We also note that any operation defined for row vectors is defined analogously for column vectors.

1.3 Vector Addition and Scalar Multiplication

Consider two vectors u and v in $\mathbb{R}^n$:

$$\boldsymbol{u} = (a_1, \ldots, a_n) \text{ and } \boldsymbol{v} = (b_1, \ldots, b_n)$$

Their *sum* can be expressed as $\boldsymbol{u} + \boldsymbol{v}$. $\boldsymbol{u} + \boldsymbol{v}$ is the *vector* that is obtained by adding the corresponding elements from $\boldsymbol{u}$ and $\boldsymbol{v}$. It is as follows:

$$\boldsymbol{u} + \boldsymbol{v} = (a_1 + b_2, \ldots, a_n + b_n)$$

The *scalar product* of the vector $\boldsymbol{u}$ by a real number k can be expressed as $k\boldsymbol{u}$. $k\boldsymbol{u}$ is the vector obtained by multiplying each element of $\boldsymbol{u}$ by k. It is as follows:

$$k\boldsymbol{u} = k(a_1, \ldots, a_n) = (ka_1, \ldots, ka_n)$$

Remark 1.2. Notice that $\boldsymbol{u} + \boldsymbol{v}$ and $k\boldsymbol{u}$ are also vectors in $\mathbb{R}^n$. The sum of vectors who have different numbers of elements is not defined.

Negatives and *subtraction* are defined in $\mathbb{R}^n$ as follows:

$$-\boldsymbol{u} = (-1)\boldsymbol{u} \qquad \text{and} \qquad \boldsymbol{u} - \boldsymbol{v} = \boldsymbol{u} + (-\boldsymbol{v})$$

The vector $-u$ is called the negative of $\boldsymbol{u}$, and $\boldsymbol{u} - \boldsymbol{v}$ is called the *difference* of $\boldsymbol{u}$ and $\boldsymbol{v}$.

Now suppose we are given vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ in $\mathbb{R}^n$ and scalars $k_1, \ldots, k_m$ in $\mathbb{R}$. We can multiply the vectors by corresponding scalars and then add the resultant scalar products to form the vector:

$$\boldsymbol{v} = k_1\boldsymbol{u}_1 + k_2\boldsymbol{u}_2 + \cdots + k_m\boldsymbol{u}_m$$

The vector $\boldsymbol{v}$ is called a *linear combination* of the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$.

Theorem 1.3 (Basic properties of vectors under the operations of vector addition and scalar multiplication)
For any vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w} \in \mathbb{R}^n$ and any scalars k, $k' \in \mathbb{R}$,

$(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$	$k(\boldsymbol{u} + \boldsymbol{v}) = k\boldsymbol{u} + k\boldsymbol{v}$
$\boldsymbol{u} + \boldsymbol{0} = \boldsymbol{u}$	$(k + k')\boldsymbol{u} = k\boldsymbol{u} + k'u$
$\boldsymbol{u} + (-\boldsymbol{u}) = \boldsymbol{0}$	$(kk')\boldsymbol{u} = k(k'\boldsymbol{u})$
$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$	$1(\boldsymbol{u}) = \boldsymbol{u}$

Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$ for which $\boldsymbol{u} = kv$ for some non-zero scalar k in $\mathbb{R}$. It follows that $\boldsymbol{u}$ is a *multiple* of $\boldsymbol{v}$. Also, $\boldsymbol{u}$ is said to be in the *same* or *opposite direction* as $\boldsymbol{v}$ for $k > 0$ or $k < 0$.

1.4 Dot Product

Consider the following arbitrary vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$:

$$\boldsymbol{u} = (a_1, \ldots, a_n) \text{ and } \boldsymbol{v} = (b_1, \ldots, b_n)$$

The *dot product* or *inner product* or *scalar product* of $\boldsymbol{u}$ and $\boldsymbol{v}$ is denoted and defined by $\boldsymbol{u} \cdot \boldsymbol{v} = a_1b_1 + \ldots a_nb_n$. $\boldsymbol{u} \cdot \boldsymbol{v}$ is obtained by multiplying corresponding components and taking the sum of the resulting products. The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *orthogonal* (or *perpendicular*) if their *dot product* is zero; $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

Theorem 1.4 (Basic properties of the dot product in $\mathbb{R}^n$)

For any vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w} \in \mathbb{R}^n$ and any scalar $k \in \mathbb{R}$,

$(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$	$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$
$(k\boldsymbol{u}) \cdot \boldsymbol{v} = k(\boldsymbol{u} \cdot \boldsymbol{v})$	$u \cdot \boldsymbol{u} \geq 0$, and $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ if $\boldsymbol{u} = 0$

To be somewhat complete (and rigorous) with the properties, consider $\boldsymbol{u} \cdot (k\boldsymbol{v})$. It is similar to the cell on row 2, column 1. It follows that $\boldsymbol{u} \cdot (k\boldsymbol{v}) = (k\boldsymbol{v}) \cdot \boldsymbol{u} = k(\boldsymbol{v} \cdot \boldsymbol{u}) = k(\boldsymbol{u} \cdot \boldsymbol{v})$.

The space $\mathbb{R}^n$ with the above operations of vector addition, scalar multiplication, and dot product is usually called **Euclidean $\boldsymbol{n}$-space**.

1.5 Norm (Length) of a Vector

The *norm* or *length* of a vector u in $\mathbb{R}^n$ can be expressed as $||\boldsymbol{u}||$. It is defined to be the non-negative square root of $\boldsymbol{u} \cdot \boldsymbol{u}$ (the inner product). For example, if $\boldsymbol{u} = (a_1, \ldots, a_n)$, then $||\boldsymbol{u}|| = \sqrt{\boldsymbol{u} \cdot \boldsymbol{u}} = \sqrt{a_1^2 + \ldots, a_n^2}$. Therefore, $||\boldsymbol{u}||$ is the square root of the sum of the squares of the elements of $\boldsymbol{u}$. Thus, $||\boldsymbol{u}|| > 0$, and $||\boldsymbol{u}|| = 0$ *iff* $\boldsymbol{u} = 0$.

A vector $\boldsymbol{u}$ is called a *unit vector* if $||\boldsymbol{u}||$ or if $\boldsymbol{u} \cdot \boldsymbol{u} = 1$ (both are equivalent).
For any non-zero vector v in $\mathbb{R}^n$, the vector $\hat{\boldsymbol{v}} = \frac{1}{||\boldsymbol{v}||}\boldsymbol{v} = \frac{\boldsymbol{v}}{||\boldsymbol{v}||}$ is the unique unit vector in the same direction as $\boldsymbol{v}$. The process of finding $\hat{\boldsymbol{v}}$ from $\boldsymbol{v}$ is called *normalizing* $\boldsymbol{v}$.

Theorem 1.5 (Cauchy-Schwarz Inequality)

For any vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$|\boldsymbol{u} \cdot \boldsymbol{v}| \leq ||\boldsymbol{u}|| \; ||\boldsymbol{v}||$$

Theorem 1.6 (Triangle Inequality or Minkowski's Inequality)

For any vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$||\boldsymbol{u} + \boldsymbol{v}|| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$$

2 Algebra of Matrices

2.1 Matrices

A matrix $\boldsymbol{A}$ over a field $\boldsymbol{K}$ is a rectangular array of scalars:

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

The rows of $\boldsymbol{A}$ are the m horizontal lists of scalars:

$$(a_{11}, \dots, a_{1n}), (a_{21}, \dots, a_{2n}), \dots, (a_{m1}, \dots, a_{mn})$$

A matrix with m rows and n columns is called an $m \times n$ matrix which is read as m by n matrix. The pair of numbers m and n is called the size of the matrix. Suppose two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are *equal*, $\boldsymbol{A} = \boldsymbol{B}$. This is true when they have the same size and if the corresponding elements are equal. Thus, the equality of two $m \times n$ matrices is equivalent to a system of mn equalities; one for each corresponding pair of elements.

A matrix with only one row or with only one column is called a *row matrix/vector* and *column matrix/vector* respectively. A matrix whose entries are all zero is called a *zero matrix* and is written as $\mathbf{0}$.

2.2 Matrix Addition and Scalar Multiplication

Let $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{B} = [b_{ij}]$ be two matrices with the same size; $m \times n$ matrices. The *sum* of $\boldsymbol{A}$ and $\boldsymbol{B}$ can be expressed as $\boldsymbol{A} + \boldsymbol{B}$. It is the matrix obtained by adding the corresponding elements from $\boldsymbol{A}$ and $\boldsymbol{B}$:

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

The *product* of a matrix $\boldsymbol{A}$ by a scalar k can be expressed as $k \cdot \boldsymbol{A} = k\boldsymbol{A}$. It is the matrix obtained by multiplying each element of $\boldsymbol{A}$ by k:

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

Negatives and *substractions* are defined in a similar fashion as vectors as seen in the first section; $-\boldsymbol{A} = (-1)\boldsymbol{A}$ and $\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{A} + (-\boldsymbol{B})$. The matrix $-\boldsymbol{A}$ is called the *negative* of the matrix $\boldsymbol{A}$. The matrix $\boldsymbol{A} - \boldsymbol{B}$ is called the *difference* of $\boldsymbol{A}$ and $\boldsymbol{B}$. The sum of matrices with different sizes is not defined.

Consider λ and μ where $\lambda, \mu \in \mathbb{R}$, and two $m \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. $\lambda\boldsymbol{A} + \mu\boldsymbol{B}$ is called a *linear combination* of $\boldsymbol{A}$ and $\boldsymbol{B}$

Theorem 2.1 (Basic properties of matrices under the operations of matrix addition and scalar multiplication)
For any matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ (with the same size) and any scalars k and k',

$(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C} = \boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})$	$k\,(\boldsymbol{A}+\boldsymbol{B}) = k\boldsymbol{A}+k\boldsymbol{B}$
$\boldsymbol{A}+\boldsymbol{0} = \boldsymbol{0}+\boldsymbol{A} = \boldsymbol{A}$	$(k+k')\boldsymbol{A} = k\boldsymbol{A}+k'\boldsymbol{A}$
$\boldsymbol{A}+(-\boldsymbol{A}) = (-\boldsymbol{A})+\boldsymbol{A} = \boldsymbol{0}$	$(kk')\boldsymbol{A} = k(k'\boldsymbol{A})$
$\boldsymbol{A}+\boldsymbol{B} = \boldsymbol{B}+\boldsymbol{A}$	$1 \cdot \boldsymbol{A} = \boldsymbol{A}$

2.3 Matrix Multiplication

Suppose $f(k)$ is an algebraic expression involving the letter k. It follows that the expression $\sum_{k=1}^{n} f(k)$ has the following meaning; $\sum$ is the *summation symbol*, and k is called the index where 1 and n are the *lower* and *upper bounds* respectively. As such, this means we are taking the sum of $f(k)$ from 1 to n. First, we set $k=1$ in $f(k)$ which yields $f(1)$. Next, we set $k=2$ in $f(k)$ which yields $f(2)$, and take sum with $f(1)$ which yields $f(1)+f(2)$. We increase the value of k by 1 until we reach n; taking the sum of the result at each step. We should obtain the sum $f(1)+f(2)+\cdots+f(n)$ at the end.

We also generalize our definition by allowing the sum to range from any integer n to N.

$$\sum_{k=n}^{N} f(k) = f(n)+f(n+1)+f(n+2)+\cdots+f(N)$$

The *product* of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be expressed as $\boldsymbol{AB}$. Matrix multiplication is quite convoluted as many things are happening but we will try to break it down slowly.
For a start, we begin with a special case. The product $\boldsymbol{AB}$ of a row matrix $\boldsymbol{A} = [a_i]$ and column matrix $\boldsymbol{B} = [b_i]$ with the same number of elements is defined to be the scalar (or 1×1 matrix) obtained by multiplying corresponding entries and adding them together:

$$\boldsymbol{AB} = [a_1, a_2, \ldots, a_n]\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \sum_{k=1}^{n} a_kb_k$$

It is important to note that $\boldsymbol{AB}$ is a scalar (or a 1×1 matrix). The product $\boldsymbol{AB}$ is not defined when $\boldsymbol{A}$ and $\boldsymbol{B}$ have different numbers of elements.

Suppose $\boldsymbol{A} = [a_{jk}]$ and $\boldsymbol{B} = b_[kj]$ are matrices such that the number of columns of $\boldsymbol{A}$ is equal to the number of rows of $\boldsymbol{B}$; Suppose $\boldsymbol{A}$ is an $m\times p$ matrix and $\boldsymbol{B}$ is a $p\times n$ matrix. Then the product $\boldsymbol{AB}$ is an $m\times n$ matrix whose ij-entry is obtained by multiplying the ith row of $\boldsymbol{A}$ by the jth column of $\boldsymbol{B}$.

$$\begin{bmatrix} a_{11} & \ldots & a_{1p} \\ \vdots & \vdots & \vdots \\ a_{i1} & \ldots & a_{ip} \\ \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mp} \end{bmatrix}\begin{bmatrix} b_{11} & \ldots & b_{1j} & \ldots & b_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{p1} & \ldots & b_{pj} & \ldots & b_{pn} \end{bmatrix} = \begin{bmatrix} c_{11} & \ldots & c_{1n} \\ \vdots & \vdots & \vdots \\ \ldots & c_{ij} & \ldots \\ \vdots & \vdots & \vdots \\ c_{m1} & \ldots & c_{mn} \end{bmatrix}$$

where $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$

The product $\boldsymbol{AB}$ is not defined if $\boldsymbol{A}$ is an $m \times p$ matrix and $\boldsymbol{B}$ is a $q \times n$ matrix, where $p \neq q$.

Example 2.2 (Schaum's Easy Outlines: Linear Algebra Crash Course - Example 2.5)

(a) Find $\boldsymbol{AB}$ where $\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 2 & 0 & -4 \\ 5 & -2 & 6 \end{bmatrix}$.

Since $\boldsymbol{A}$ is 2×2 and $\boldsymbol{B}$ is 2×3, the product $\boldsymbol{AB}$ is defined and $\boldsymbol{AB}$ is a 2×3 matrix. To obtain the first row of $\boldsymbol{AB}$, multiply the first row of $\boldsymbol{A}$ by each column of $\boldsymbol{B}$. Then multiply the second row of $\boldsymbol{A}$ by each column of $\boldsymbol{B}$.

$$\boldsymbol{AB} = \begin{bmatrix} 2+15 & 0-6 & -4+18 \\ 4-5 & 0+2 & -8-6 \end{bmatrix} = \begin{bmatrix} 17 & -6 & 14 \\ 4-5 & 2 & -14 \end{bmatrix}$$

(b) Suppose $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix}$.

$$\boldsymbol{AB} = \begin{bmatrix} 5+0 & 6-4 \\ 15+0 & 18-8 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 15 & 10 \end{bmatrix}$$

$$\boldsymbol{BA} = \begin{bmatrix} 5+18 & 10+24 \\ 0-6 & 0-8 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ -6 & -8 \end{bmatrix}$$

In part (b) of the example above, it shows that matrix multiplication is not commutative; the products $\boldsymbol{AB}$ and $\boldsymbol{BA}$ of matrices need not be equal.

Theorem 2.3 (Basic properties that matrix multiplication does satisfy)

- $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$ (associative law),
- $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$ (left distributive law),
- $(\boldsymbol{B}+\boldsymbol{C})\boldsymbol{A} = \boldsymbol{BA} + \boldsymbol{CA}$ (right distributive law),
- $k(\boldsymbol{AB}) = (k\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}k\boldsymbol{B}$, where k is a scalar.

We note that $\boldsymbol{0A} = \boldsymbol{0}$ and $\boldsymbol{B0} = \boldsymbol{0}$, where $\boldsymbol{0}$ is the zero matrix.

2.4 Transpose of a Matrix

The *transpose* of matrix $\boldsymbol{A}$ can be expressed as $\boldsymbol{A}^T$. It is the matrix obtained by writing the columns of $\boldsymbol{A}$, in order, as rows. Consider the following,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & -3 & -5 \end{bmatrix}^T = \begin{bmatrix} 1 \\ -3 \\ -5 \end{bmatrix}$$

In other words, if $\boldsymbol{A} = [a_{ij}]$ is an $m \times n$ matrix, then $\boldsymbol{A}^T = [b_{ij}]$ is the $n \times m$ matrix where $b_{ij} = a_{ij}$.

Theorem 2.4 (Basic properties of the transpose operation)
Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices and let k be a scalar. Then, whenever the sum and product are defined:

$(\boldsymbol{A}+\boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$	$(k\boldsymbol{A})^T = k\boldsymbol{A}^T$
$(\boldsymbol{A}^T)^T = \boldsymbol{A}$	$(\boldsymbol{A}\boldsymbol{B})^T = \boldsymbol{B}^T\boldsymbol{A}^T$

In the last cell – row 2, column 2 – we see that the transpose of a product is the product of the transposes, but in the reverse order.

2.5 Square Matrices

A *square matrix* is a matrix with the same number of rows as columns. An $n \times n$ square matrix is said to be of *order* n and is sometimes called an n*-square matrix*.

Recall that not every two matrices can be added or multiplied. However, if we only consider square matrices of some given order n, then this inconvenience disappears. Specifically, the operations of addition, multiplication, sclaar multiplication, and transpose can be performed on any $n \times n$ matrices, and the result is again an $n \times n$ matrix.

2.6 Diagonal and Trace

Let $\boldsymbol{A} = [a_{ij}]$ be an n-square matrix. The *diagonal* or *main diagonal* of $\boldsymbol{A}$ consists of the elements with the same subscripts:

$$a_{11}, a_{22}, a_{33}, \ldots, a_{nn}$$

The *trace* of $\boldsymbol{A}$ can be expressed as $\mathrm{tr}(\boldsymbol{A})$. It is the sum of the diagonal elements; $\mathrm{tr}(\boldsymbol{A}) = a_{11} + a_{22} + a_{33} + \cdots + a_{nn}$.

Theorem 2.5 (Basic properties of the trace of a matrix)
Suppose $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{B} = [b_{ij}]$ are n-square matrices and k is a scalar.

$\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$	$\operatorname{tr}\left(\boldsymbol{A}^T\right) = \operatorname{tr}(\boldsymbol{A})$
$\operatorname{tr}(k\boldsymbol{A}) = k\operatorname{tr}(\boldsymbol{A})$	$\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$

Although $\boldsymbol{AB} \neq \boldsymbol{BA}$, the traces are equal.

2.7 Identity Matrix, Scalar Matrices

The n-square *identity* or *unit matrix*, dentoed by $\boldsymbol{I}_n$, or simply $\boldsymbol{I}$, is the n-square matrix with 1s on thee diagonal and 0s everywhere. The identity matrix $\boldsymbol{I}$ is similar to the scalar 1 in that, for any n-square matrix $\boldsymbol{A}$, $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$. More generally, if $\boldsymbol{B}$ is an $m \times n$ matrix:

$$\boldsymbol{BI_n} = \boldsymbol{I_m B} = \boldsymbol{B}$$

For any scalar k, the matrix $k\boldsymbol{I}$ that contains ks on the diagonal and 0s everywhere else is called the scalar matrix corresponding to the scalar k. Multiplying a matrix $\boldsymbol{A}$ by the scalar matrix $k\boldsymbol{I}$ is equivalent to multiplying $\boldsymbol{A}$ by the scalar k; $(k\boldsymbol{I})\boldsymbol{A} = k(\boldsymbol{IA}) = k\boldsymbol{A}$.

The *Kronecker delta* function δ_{ij} can also be used to define the identity matrix;

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

Therefore, the identity matrix, $\boldsymbol{I} = [\delta_{ij}]$.

2.8 Polynomials in Matrices

Let $\boldsymbol{A}$ be an n-square matrix over a field K. *Powers* of $\boldsymbol{A}$:

$$\boldsymbol{A}^2 = \boldsymbol{AA}, \quad \boldsymbol{A}^3 = \boldsymbol{A}^2\boldsymbol{A}, \quad \ldots, \quad \boldsymbol{A}^{n+1} = \boldsymbol{A}^n\boldsymbol{A}, \quad \ldots, \quad \text{and} \quad \boldsymbol{A}^0 = \boldsymbol{I}$$

Polynomials in the matrix $\boldsymbol{A}$ are also defined. For any polynomial $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ where a_i are scalars in K, $f(\boldsymbol{A})$ is defined to be the following matrix:

$$f(\boldsymbol{A}) = a_0\boldsymbol{I} + a_1\boldsymbol{A} + a_2\boldsymbol{A}^2 + \cdots + a_n\boldsymbol{A}^n$$

If $f(\boldsymbol{A})$ is the zero matrix, then $\boldsymbol{A}$ is called a *zero* or *root* of $f(x)$.

2.9 Invertible (Non-singular) Matrices

A square matrix $\boldsymbol{A}$ is said to be *invertible* or *non-singular* if there exists a matrix $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix. If $\boldsymbol{AB_1} = \boldsymbol{B_1A} = \boldsymbol{I}$ and $\boldsymbol{AB_2} = \boldsymbol{B_2A} = \boldsymbol{I}$, then $\boldsymbol{B_1} = \boldsymbol{B_1I} = \boldsymbol{B_1(AB_2)} = \boldsymbol{(B_1A)B_2} = \boldsymbol{IB_2} = \boldsymbol{B_2}$.

We call such a matrix $\boldsymbol{B}$ the *inverse* of $\boldsymbol{A}$. It can be expressed as $\boldsymbol{A}^{-1}$. Do note that the above relation is symmetric; if $\boldsymbol{B}$ is the inverse of $\boldsymbol{A}$, then $\boldsymbol{A}$ is the inverse of $\boldsymbol{B}$.

It is known that $\boldsymbol{AB} = \boldsymbol{I}$ *iff* $\boldsymbol{BA} = \boldsymbol{I}$. Therefore it is sufficient to test only one product to determine whether or not two given matrices are inverses.

Now suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible. It follows that $\boldsymbol{AB}$ is invertible and $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$. More generally, if $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_k$ are invertible, then their product is invertible, and $(\boldsymbol{A}_1\ \boldsymbol{A}_2\ \ldots\ \boldsymbol{A}_k)^{-1} = \boldsymbol{A}_k^{-1}\ \ldots\ \boldsymbol{A}_2^{-1}\ \boldsymbol{A}_1^{-1}$, the product of the inverses in the reverse order.

2.10 Inverse of a 2×2 Matrix

Let $\boldsymbol{A}$ be an arbitrary 2×2 matrix; $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. We want to derive a formula for $\boldsymbol{A}^{-1}$. In this case, we seek $2^2 = 4$ scalars – x_1, y_1, x_2, y_2 – such that:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} ax_1 + by_1 & ax_2 + by_2 \\ cx_1 + dy_1 & cx_2 + dy_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Setting the four entries equal to the corresponding entries in the identity matrix yields four equations, which can be partitioned into 2×2 systems as follows:

$$\begin{aligned} ax_1 + by_1 &= 1, \quad & ax_2 + by_2 &= 0 \\ cx_1 + dy_1 &= 0, \quad & cx_2 + dy_2 &= 1 \end{aligned}$$

Suppose we let $|\boldsymbol{A}| = ad - bc$ ($|\boldsymbol{A}|$ is the *determinant* of $\boldsymbol{A}$). Assuming $|\boldsymbol{A}| \neq 0$ (if $|\boldsymbol{A}| = 0$, $\boldsymbol{A}$ is not invertible), we can solve uniquely for the above unknowns x_1, y_1, x_2, y_2 which obtains:

$$x_1 = \frac{d}{|\boldsymbol{A}|}, \quad y_1 = \frac{-c}{|\boldsymbol{A}|}, \quad x_2 = \frac{-b}{|\boldsymbol{A}|}, \quad y_2 = \frac{a}{|\boldsymbol{A}|}$$

Accordingly,

$$\boldsymbol{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} d/|\boldsymbol{A}| & -b/|\boldsymbol{A}| \\ -c/|\boldsymbol{A}| & a/|\boldsymbol{A}| \end{bmatrix} = \frac{1}{|\boldsymbol{A}|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

In other words, when $|\boldsymbol{A}| \neq 0$, the inverse of a 2×2 matrix $\boldsymbol{A}$; $\boldsymbol{A}^{-1}$ is obtained by:

1. Interchange the two elements on the diagonal,
2. Take the negatives of the other two elements,
3. Multiply the reuslting matrix by $1/|\boldsymbol{A}|$ or divide each element by $|\boldsymbol{A}|$.

To find the inverse of a matrix in general, it is finding the solution of a collection of $n \times n$ systems of linear equations for an arbitrary n-square matrix.

2.11 Special Types of Square Matrices

2.11.1 Diagonal Matrices

As square matrix $\boldsymbol{D} = [d_{ij}]$ is *diagonal* if its non-diagonal entries are all zero. It can be expressed as:

$$\boldsymbol{D} = \text{diag}(d_{11}, d_{22}, \dots, d_{nn})$$

where some or all the d_{ii} may be zero.

2.11.2 Triangular Matrices

A square matrix is *upper triangular* or simply *triangular* if all entries below the main diagonal are equal to 0; If $a_{ij} = 0$ for all $i > j$.

A *lower triangular* matrix is a square matrix whose entries above the diagonal are all zero.

Theorem 2.6

Suppose $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{B} = [b_{ij}]$ are $n \times n$ upper/lower triangular matrices.

1. $\boldsymbol{A} + \boldsymbol{B}, k\boldsymbol{A}, \boldsymbol{AB}$ are triangular with respective diagonals.
   $(a_{11} + b_{11}, \dots, a_{nn} + b_{nn})$, $(ka_{11}, \dots, ka_{nn})$, $(a_{11}b_{11}, \dots, a_{nn}b_{nn})$

2. For any polynomial $f(x)$, the matrix $f(\boldsymbol{A})$ is triangular with diagonal $(f(a_{11}), f(a_{22}), \dots, f(a_{nn}))$

3. $\boldsymbol{A}$ is invertible *iff* each diagonal element $a_{ii} \neq 0$, and when $\boldsymbol{A}^{-1}$ exists it is also triangular.

2.11.3 Symmetric Matrices

A matrix $\boldsymbol{A}$ is *symmetric* if $\boldsymbol{A}^T = \boldsymbol{A}$. Equivalently, $\boldsymbol{A} = [a_{ij}]$ is symmetric if *symmetric elements* (mirror elements wrt. the diagonal) are equal; if each $a_{ij} = a_{ji}$.

A matrix $\boldsymbol{A}$ is *skew-symmetric* if $\boldsymbol{A}^T = -\boldsymbol{A}$. Equivalently, $a_{ij} = -a_{ji}$. Clearly the diagonal elements of such am atrix must be zero, since $a_{ii} = -a_{ii} \Rightarrow a_{ii} = 0$.

2.11.4 Orthogonal Matrices

A real matrix $\boldsymbol{A}$ is *orthogonal* if $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$; $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$. Therefore, $\boldsymbol{A}$ must necessarily be square and invertible.

Suppose $\boldsymbol{A}$ is a real orthogonal 3×3 matrix with rows:

$$u_1 = (a_1, a_2, a_3), \quad u_2 = (b_1, b_2, b_3), \quad u_3 = (c_1, c_2, c_3).$$

Since $\boldsymbol{A}$ is orthogonal, we must have $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}$.

$$\boldsymbol{A}\boldsymbol{A}^T = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{I}$$

Multiplying $\boldsymbol{A}$ by $\boldsymbol{A}^T$ and setting each entry equal to the corresponding entry in $\boldsymbol{I}$ yields a 3×3 system of equations.

$$\begin{aligned} a_1^2 + a_2^2 + a_3^3 = 1, \quad & a_1b_1 + a_2b_2 + a_3b_3 = 0, \quad & a_1c_1 + a_2c_2 + a_3c_3 = 0 \\ b_1a_1 + b_2a_2 + b_3a_3 = 0, \quad & b_1^2 + b_2^2 + b_3^2 = 1, \quad & b_1c_1 + b_2c_2 + b_3c_3 = 0 \\ c_1a_1 + c_2a_2 + c_3a_3 = 0, \quad & c_1b_1 + c_2b_2 + c_3b_3 = 0, \quad & c_1^2 + c_2^2 + c_3^2 = 1 \end{aligned}$$

This shows that $u_1 \cdot u_1 = 1, u_2 \cdot u_2 = 1, u_3 \cdot u_3 = 1$, and $u_i \cdot u_j = 0$ for all $i \neq j$. It follows that the rows u_1, u_2, u_3 are unit vectors and are orthogonal to each other.

In general, vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m$ in $\mathbb{R}^n$ are said to form an *orthonormal* set of vectors if the vectors are unit vectors and are orthogonal to each other:

$$u_i \cdot u_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

You can see that $u_i \cdot u_j = \delta_{ij}$ where δ_{ij} is the Kronecker delta function.

We have shown that the condition $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}$ implies that the rows of $\boldsymbol{A}$ form an orthonormal set of vectors. The condition $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$ similary implies that the columns of $\boldsymbol{A}$ also form an orthonormal set of vectors. Moreover, since each step is reversible, the converse is true. The above results for 3×3 matrices is true in general.

Theorem 2.7

Let $\boldsymbol{A}$ be a real matrix. Then the following are equivalent:

1. $\boldsymbol{A}$ is orthogonal,
2. The rows of $\boldsymbol{A}$ form an orthonormal set,
3. The columns of $\boldsymbol{A}$ form an orthonormal set.

2.11.5 Normal Vectors

A real matrix $\boldsymbol{A}$ is *norma* if it *commutes* with its transpose $\boldsymbol{A}^T$; $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{A}^T\boldsymbol{A}$. If $\boldsymbol{A}$ is symmetric, orthogonal, or skew-symmetric, then $\boldsymbol{A}$ is normal.

2.12 Block Matrices

Using a system of horizontal and vertical lines, we can partition a matrix $\boldsymbol{A}$ into sub-matrices called *blocks/cells* of $\boldsymbol{A}$. Any given matrix may be divided into blocks in different ways.

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

The convenience of block matrices; $\boldsymbol{A}$ and $\boldsymbol{B}$ into blocks is that the result of operations on $\boldsymbol{A}$ and $\boldsymbol{B}$ can be obtained by carrying out the computation with the blocks, just as if they were the actual elements of the matrices.

Suppose that $\boldsymbol{A} = [\boldsymbol{A}_{ij}]$ and $\boldsymbol{B} = [\boldsymbol{B}_{ij}]$ are block matrices with the same number of row and column blocks. Suppose that the corresponding blocks have the same size. It follows that adding the corresponding blocks of $\boldsymbol{A}$ and $\boldsymbol{B}$ also adds the corresponding elements of $\boldsymbol{A}$ and $\boldsymbol{B}$ and multiplying each block of $\boldsymbol{A}$ by a scalar k multiplies each element of $\boldsymbol{A}$ by k.

$$\boldsymbol{A}+\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{11}+\boldsymbol{B}_{11} & \boldsymbol{A}_{12}+\boldsymbol{B}_{12} & \dots & \boldsymbol{A}_{1n}+\boldsymbol{B}_{1n} \\ \boldsymbol{A}_{21}+\boldsymbol{B}_{21} & \boldsymbol{A}_{22}+\boldsymbol{B}_{22} & \dots & \boldsymbol{A}_{2n}+\boldsymbol{B}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{m1}+\boldsymbol{B}_{m1} & \boldsymbol{A}_{m2}+\boldsymbol{B}_{m2} & \dots & \boldsymbol{A}_{mn}+\boldsymbol{B}_{mn} \end{bmatrix}$$

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

The case of matrix multiplication is less obvious. Suppose $\boldsymbol{U} = [\boldsymbol{U}_{ik}]$ and $\boldsymbol{V} = [\boldsymbol{V}_{kj}]$ are block matrices such that the number of clumns of each block $\boldsymbol{U}_{ik}$ is equal to the number of rows of each block $\boldsymbol{V}_{kj}$; the condition for matrix multiplication has been satisfied.

$$\boldsymbol{UV} = \begin{bmatrix} \boldsymbol{W}_{11} & \boldsymbol{W}_{12} & \dots & \boldsymbol{W}_{1n} \\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22} & \dots & \boldsymbol{W}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{W}_{m1} & \boldsymbol{W}_{m2} & \dots & \boldsymbol{W}_{mn} \end{bmatrix}$$

where $\boldsymbol{W}_{ij} = \boldsymbol{U}_{i1}\boldsymbol{V}_{1j} + \boldsymbol{U}_{i2}\boldsymbol{V}_{2j} + \dots + \boldsymbol{U}_{ip}\boldsymbol{V}_{pj}$.

2.13 Square Block Matrices

Theorem 2.8

Let $\boldsymbol{M}$ be a block matrix. For $\boldsymbol{M}$ to be a *square block matrix*:

1. $\boldsymbol{M}$ is square matrix,
2. The blocks form a square matrix,
3. The diagonal blocks are also square matrices.

The last two conditions will occur *iff* there are the same number of horizontal and vertical lines and they are placed symmetrically.

$$\boldsymbol{A} = \left[\begin{array}{cc|cc|c} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \\ \hline 3 & 1 & 4 & 5 & 9 \\ \hline 4 & 4 & 4 & 4 & 4 \\ 3 & 5 & 3 & 5 & 3 \end{array}\right] \text{ and } \boldsymbol{B} = \left[\begin{array}{cc|cc|c} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \\ \hline 3 & 1 & 4 & 5 & 9 \\ 4 & 4 & 4 & 4 & 4 \\ \hline 3 & 5 & 3 & 5 & 3 \end{array}\right]$$

The block matrix $\boldsymbol{A}$ is not a square block matrix, since the second and third diagonal blocks are not square. On the other hand, the block matrix $\boldsymbol{B}$ is a square block matrix.

2.14 Block Diagonal Matrices

Let $\boldsymbol{M} = [\boldsymbol{A}_{ij}]$ be a square block matrix such that the non-diagonal blocks are all zero matrices; $\boldsymbol{A}_{ij} = 0$ when $i \neq j$. Then $\boldsymbol{M}$ is called a *block diagonal matrix*. It can be expressed as:

$$\boldsymbol{M} = \operatorname{diag}(\boldsymbol{A}_{11}, \boldsymbol{A}_{22}, \ldots, \boldsymbol{A}_{rr}) \text{ or } \boldsymbol{M} = \boldsymbol{A}_{11} \oplus \boldsymbol{A}_{22} \oplus \cdots \oplus \boldsymbol{A}_{rr}$$

The importance of block diagonal matrices is that the algebra of the block matrix is frequently reduced to the algebra of the individual blocks. Specifically, suppose $f(x)$ is a polynomial and $\boldsymbol{M}$ is the above block diagonal matrix. It follows that $f(\boldsymbol{M})$ is a block diagonal matrix and

$$f(\boldsymbol{M}) = \operatorname{diag}(f(\boldsymbol{A}_{11}), f(\boldsymbol{A}_{22}), \ldots, f(\boldsymbol{A}_{rr}))$$

Moreover, $\boldsymbol{M}$ is invertible *iff* each $\boldsymbol{A}_{ii}$ is invertible, and in this case, $\boldsymbol{M}^{-1}$ is a block diagonal matrix:

$$\boldsymbol{M}^{-1} = \operatorname{diag}\big(\boldsymbol{A}_{11}^{-1}, \boldsymbol{A}_{22}^{-1}, \ldots, \boldsymbol{A}_{rr}^{-1}\big)$$

3 Systems of Linear Equations

4 Vector Spaces

5 Inner Product Spaces; Orthogonality

6 Determinants

7 Diagonalization; Eigenvalues and Eigenvectors

8 Linear Mappings

9 Canonical Forms