Contents

Lecture 1

Systems of Linear Equations

In this lecture, we will introduce linear systems and the method of row reduction to solve them. We will introduce matrices as a convenient structure to represent and solve linear systems. Lastly, we will discuss geometric interpretations of the solution set of a linear system in 2- and 3-dimensions.

1.1 What is a system of linear equations?

Definition 1.1: A **system of m linear equations** in n unknown variables $x_1, x_2, \ldots, x_n$ is a collection of m equations of the form

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{1.1}$$

The numbers a_{ij} are called the **coefficients** of the linear system; because there are m equations and n unknown variables there are thefore $m \times n$ coefficients. The main problem with a linear system is of course to solve it:

Problem: Find a list of n numbers $(s_1, s_2, \ldots, s_n)$ that satisfy the system of linear equations (1.1).

In other words, if we substitute the list of numbers $(s_1, s_2, \ldots, s_n)$ for the unknown variables $(x_1, x_2, \ldots, x_n)$ in equation (1.1) then the left-hand side of the ith equation will equal b_i. We call such a list $(s_1, s_2, \ldots, s_n)$ **a solution** to the system of equations. Notice that we say "a solution" because there may be more than one. The **set** of all solutions to a linear system is called its **solution set**. As an example of a linear system, below is a linear system consisting of $m = 2$ equations and $n = 3$ unknowns:

$$\begin{aligned} x_1 - 5x_2 - 7x_3 &= 0 \\ 5x_2 + 11x_3 &= 1 \end{aligned}$$

Here is a linear system consisting of $m = 3$ equations and $n = 2$ unknowns:

$$\begin{aligned} -5x_1 + x_2 &= -1 \\ \pi x_1 - 5x_2 &= 0 \\ 63x_1 - \sqrt{2}x_2 &= -7 \end{aligned}$$

And finally, below is a linear system consisting of $m = 4$ equations and $n = 6$ unknowns:

$$\begin{aligned} -5x_1 + x_3 - 44x_4 - 55x_6 &= -1 \\ \pi x_1 - 5x_2 - x_3 + 4x_4 - 5x_5 + \sqrt{5}x_6 &= 0 \\ 63x_1 - \sqrt{2}x_2 - \tfrac{1}{5}x_3 + \ln(3)x_4 + 4x_5 - \tfrac{1}{33}x_6 &= 0 \\ 63x_1 - \sqrt{2}x_2 - \frac{1}{5}x_3 - \tfrac{1}{8}x_4 - 5x_6 &= 5 \end{aligned}$$

Example 1.2. Verify that $(1, 2, -4)$ is a solution to the system of equations

$$\begin{aligned} 2x_1 + 2x_2 + x_3 &= 2 \\ x_1 + 3x_2 - x_3 &= 11. \end{aligned}$$

Is $(1, -1, 2)$ a solution to the system?

Solution. The number of equations is $m = 2$ and the number of unknowns is $n = 3$. There are $m \times n = 6$ coefficients: $a_{11} = 2$, $a_{12} = 1$, $a_{13} = 1$, $a_{21} = 1$, $a_{22} = 3$, and $a_{23} = -1$. And $b_1 = 0$ and $b_2 = 11$. The list of numbers $(1, 2, -4)$ is a solution because

$$2 \cdot (1) + 2(2) + (-4) = 2$$

$$(1) + 3 \cdot (2) - (-4) = 11$$

On the other hand, for $(1, -1, 2)$ we have that

$$2(1) + 2(-1) + (2) = 2$$

but

$$1 + 3(-1) - 2 = -4 \neq 11.$$

Thus, $(1, -1, 2)$ is not a solution to the system. □

A linear system may not have a solution at all. If this is the case, we say that the linear system is **inconsistent**:

www.RanaMaths.com

INCONSISTENT ⇔ NO SOLUTION

A linear system is called **consistent** if it has at least one solution:

CONSISTENT ⇔ AT LEAST ONE SOLUTION

We will see shortly that a consistent linear system will have either just one solution or infinitely many solutions. For example, a linear system cannot have just 4 or 5 solutions. If it has multiple solutions, then it will have infinitely many solutions.

Example 1.3. Show that the linear system does not have a solution.

$$\begin{aligned} -x_1 + x_2 &= 3 \\ x_1 - x_2 &= 1. \end{aligned}$$

Solution. If we add the two equations we get

$$0 = 4$$

which is a contradiction. Therefore, there does not exist a list (s_1, s_2) that satisfies the system because this would lead to the contradiction $0 = 4$. □

Example 1.4. Let t be an arbitrary real number and let

$$\begin{aligned} s_1 &= -\tfrac{3}{2} - 2t \\ s_2 &= \tfrac{3}{2} + t \\ s_3 &= t. \end{aligned}$$

Show that for any choice of the parameter t, the list (s_1, s_2, s_3) is a solution to the linear system

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

Solution. Substitute the list (s_1, s_2, s_3) into the left-hand-side of the first equation

$$\left(-\tfrac{3}{2} - 2t\right) + \left(\tfrac{3}{2} + t\right) + t = 0$$

and in the second equation

$$\left(-\tfrac{3}{2} - 2t\right) + 3(\tfrac{3}{2} + t) - t = -\tfrac{3}{2} + \tfrac{9}{2} = 3$$

Both equations are satisfied for any value of t. Because we can vary t arbitrarily, we get an infinite number of solutions parameterized by t. For example, compute the list (s_1, s_2, s_3) for $t = 3$ and confirm that the resulting list is a solution to the linear system. □

www.RanaMaths.com

1.2 Matrices

We will use **matrices** to develop systematic methods to solve linear systems and to study the properties of the solution set of a linear system. Informally speaking, a **matrix** is an array or table consisting of *rows* and *columns*. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 7 & 11 & -5 \end{bmatrix}$$

is a matrix having $m = 3$ rows and $n = 4$ columns. In general, a matrix with m rows and n columns is a $m \times n$ matrix and the set of all such matrices will be denoted by $M_{m\times n}$. Hence, $\mathbf{A}$ above is a 3×4 matrix. The entry of $\mathbf{A}$ in the ith row and jth column will be denoted by a_{ij}. A matrix containing only one column is called a **column vector** and a matrix containing only one row is called a **row vector**. For example, here is a row vector

$$\mathbf{u} = \begin{bmatrix} 1 & -3 & 4 \end{bmatrix}$$

and here is a column vector

$$\mathbf{v} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

We can associate to a linear system three matrices: (1) the coefficient matrix, (2) the output column vector, and (3) the augmented matrix. For example, for the linear system

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

the coefficient matrix $\mathbf{A}$, the output vector $\mathbf{b}$, and the augmented matrix $[\mathbf{A}\ \mathbf{b}]$ are:

$$\mathbf{A} = \begin{bmatrix} 5 & -3 & 8 \\ 1 & 4 & -6 \\ 0 & 2 & 4 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}, \quad [\mathbf{A}\ \mathbf{b}] = \begin{bmatrix} 5 & -3 & 8 & -1 \\ 1 & 4 & -6 & 0 \\ 0 & 2 & 4 & 3 \end{bmatrix}.$$

If a linear system has m equations and n unknowns then the coefficient matrix $\mathbf{A}$ must be a $m \times n$ matrix, that is, $\mathbf{A}$ has m rows and n columns. Using our previously defined notation, we can write this as $\mathbf{A} \in M_{m\times n}$.

If we are given an augmented matrix, we can write down the associated linear system in an obvious way. For example, the linear system associated to the augmented matrix

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

is

$$\begin{aligned} x_1 + 4x_2 - 2x_3 + 8x_4 &= 12 \\ x_2 - 7x_3 + 2x_4 &= -4 \\ 5x_3 - x_4 &= 7. \end{aligned}$$

www.RanaMaths.com

We can study matrices without interpreting them as coefficient matrices or augmented matrices associated to a linear system. **Matrix algebra** is a fascinating subject with numerous applications in every branch of engineering, medicine, statistics, mathematics, finance, biology, chemistry, etc.

1.3 Solving linear systems

In algebra, you learned to solve equations by first "simplifying" them using operations that do not alter the solution set. For example, to solve $2x = 8 - 2x$ we can add to both sides $2x$ and obtain $4x = 8$ and then multiply both sides by $\frac{1}{4}$ yielding $x = 2$. We can do similar operations on a linear system. There are three basic operations, called **elementary operations**, that can be performed:

1. Interchange two equations.
2. Multiply an equation by a nonzero constant.
3. Add a multiple of one equation to another.

These operations do not alter the solution set. The idea is to apply these operations iteratively to simplify the linear system to a point where one can easily write down the solution set. It is convenient to apply elementary operations on the augmented matrix $[\mathbf{A}\ \mathbf{b}]$ representing the linear system. In this case, we call the operations **elementary row operations**, and the process of simplifying the linear system using these operations is called **row reduction**. The goal with row reducing is to transform the original linear system into one having a **triangular structure** and then perform **back substitution** to solve the system. This is best explained via an example.

Example 1.5. Use back substitution on the augmented matrix

$$\begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

to solve the associated linear system.

Solution. Notice that the augmented matrix has a triangular structure. The third row corresponds to the equation $x_3 = 1$. The second row corresponds to the equation

$$x_2 - x_3 = 0$$

and therefore $x_2 = x_3 = 1$. The first row corresponds to the equation

$$x_1 - 2x_3 = -4$$

and therefore

$$x_1 = -4 + 2x_3 = -4 + 2 = -2.$$

Therefore, the solution is $(-2, 1, 1)$. □

www.RanaMaths.com

Example 1.6. Solve the linear system using elementary row operations.

$$\begin{aligned} -3x_1 + 2x_2 + 4x_3 &= 12 \\ x_1 - 2x_3 &= -4 \\ 2x_1 - 3x_2 + 4x_3 &= -3 \end{aligned}$$

Solution. Our goal is to perform elementary row operations to obtain a triangular structure and then use back substitution to solve. The augmented matrix is

$$\begin{bmatrix} -3 & 2 & 4 & 12 \\ 1 & 0 & -2 & -4 \\ 2 & -3 & 4 & -3 \end{bmatrix}.$$

Interchange Row 1 (R_1) and Row 2 (R_2):

$$\begin{bmatrix} -3 & 2 & 4 & 12 \\ 1 & 0 & -2 & -4 \\ 2 & -3 & 4 & -3 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 0 & -2 & -4 \\ -3 & 2 & 4 & 12 \\ 2 & -3 & 4 & -3 \end{bmatrix}$$

As you will see, this first operation will simplify the next step. Add $3R_1$ to R_2:

$$\begin{bmatrix} 1 & 0 & -2 & -4 \\ -3 & 2 & 4 & 12 \\ 2 & -3 & 4 & -3 \end{bmatrix} \xrightarrow{3R_1 + R_2} \begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 2 & -2 & 0 \\ 2 & -3 & 4 & -3 \end{bmatrix}$$

Add $-2R_1$ to R_3:

$$\begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 2 & -2 & 0 \\ 2 & -3 & 4 & -3 \end{bmatrix} \xrightarrow{-2R_1 + R_3} \begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 2 & -2 & 0 \\ 0 & -3 & 8 & 5 \end{bmatrix}$$

Multiply R_2 by $\frac{1}{2}$:

$$\begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 2 & -2 & 0 \\ 0 & -3 & 8 & 5 \end{bmatrix} \xrightarrow{\frac{1}{2}R_2} \begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 1 & -1 & 0 \\ 0 & -3 & 8 & 5 \end{bmatrix}$$

Add $3R_2$ to R_3:

$$\begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 1 & -1 & 0 \\ 0 & -3 & 8 & 5 \end{bmatrix} \xrightarrow{3R_2 + R_3} \begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 5 & 5 \end{bmatrix}$$

Multiply R_3 by $\frac{1}{5}$:

$$\begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 5 & 5 \end{bmatrix} \xrightarrow{\frac{1}{5}R_3} \begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

We can continue row reducing but the row reduced augmented matrix is in triangular form. So now use back substitution to solve. The linear system associated to the row reduced

www.RanaMaths.com

augmented matrix is

$$\begin{aligned} x_1 - 2x_3 &= -4 \\ x_2 - x_3 &= 0 \\ x_3 &= 1 \end{aligned}$$

The last equation gives that $x_3 = 1$. From the second equation we obtain that $x_2 - x_3 = 0$, and thus $x_2 = 1$. The first equation then gives that $x_1 = -4 + 2(1) = -2$. Thus, the solution to the original system is $(-2, 1, 1)$. You should verify that $(-2, 1, 1)$ is a solution to the original system. □

The original augmented matrix of the previous example is

$$\mathbf{M} = \begin{bmatrix} -3 & 2 & 4 & 12 \\ 1 & 0 & -2 & -4 \\ 2 & -3 & 4 & -3 \end{bmatrix} \quad \rightarrow \quad \begin{aligned} -3x_1 + 2x_2 + 4x_3 &= 12 \\ x_1 - 2x_3 &= -4 \\ 2x_1 - 3x_2 + 4x_3 &= -3. \end{aligned}$$

After row reducing we obtained the row reduced matrix

$$\mathbf{N} = \begin{bmatrix} 1 & 0 & -2 & -4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \rightarrow \quad \begin{aligned} x_1 - 2x_3 &= -4 \\ x_2 - x_3 &= 0 \\ x_3 &= 1. \end{aligned}$$

Although the two augmented matrices $\mathbf{M}$ and $\mathbf{N}$ are clearly distinct, it is a fact that they have the same solution set.

Example 1.7. Using elementary row operations, show that the linear system is inconsistent.

$$\begin{aligned} x_1 + 2x_3 &= 1 \\ x_2 + x_3 &= 0 \\ 2x_1 + 4x_3 &= 1 \end{aligned}$$

Solution. The augmented matrix is

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 4 & 1 \end{bmatrix}$$

Perform the operation $-2R_1 + R_3$:

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 4 & 1 \end{bmatrix} \xrightarrow{-2R_1+R_3} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

The last row of the simplified augmented matrix

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

www.RanaMaths.com

corresponds to the equation

$$0x_1 + 0x_2 + 0x_3 = -1$$

Obviously, there are no numbers x_1, x_2, x_3 that satisfy this equation, and therefore, the linear system is inconsistent, i.e., it has no solution. In general, if we obtain a row in an **augmented matrix** of the form

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & c \end{bmatrix}$$

where c is a **nonzero** number, then the linear system is inconsistent. We will call this type of row an **inconsistent row**. However, a row of the form

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

corresponds to the equation $x_2 = 0$ which is perfectly valid. □

1.4 Geometric interpretation of the solution set

The set of points (x_1, x_2) that satisfy the linear system

$$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 3x_2 &= 3 \end{aligned} \tag{1.2}$$

is the intersection of the two lines determined by the equations of the system. The solution for this system is $(3, 2)$. The two lines intersect at the point $(x_1, x_2) = (3, 2)$, see Figure 1.1.

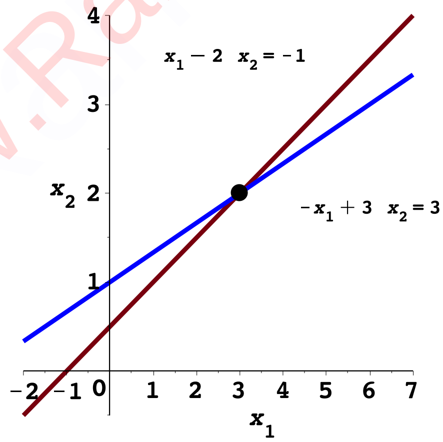


Figure 1.1: The intersection point of the two lines is the solution of the linear system (1.2)

Similarly, the solution of the linear system

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ -4x_1 + 5x_2 + 9x_3 &= -9 \end{aligned} \tag{1.3}$$

www.RanaMaths.com

is the intersection of the three planes determined by the equations of the system. In this case, there is only one solution: $(29, 16, 3)$. In the case of a consistent system of two equations, the solution set is the line of intersection of the two planes determined by the equations of the system, see Figure 1.2.

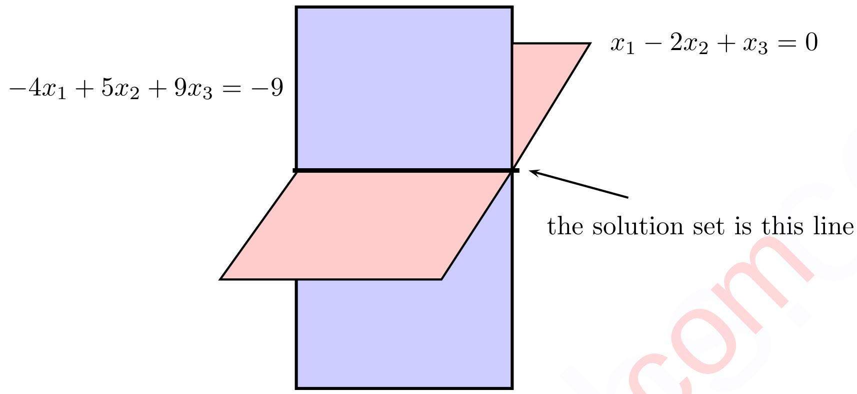


Figure 1.2: The intersection of the two planes is the solution set of the linear system (1.3)

After this lecture you should know the following:

- what a linear system is
- what it means for a linear system to be consistent and inconsistent
- what matrices are
- what are the matrices associated to a linear system
- what the elementary row operations are and how to apply them to simplify a linear system
- what it means for two matrices to be row equivalent
- how to use the method of back substitution to solve a linear system
- what an inconsistent row is
- how to identify using elementary row operations when a linear system is inconsistent
- the geometric interpretation of the solution set of a linear system

www.RanaMaths.com

www.RanaMaths.com

Lecture 2

Row Reduction and Echelon Forms

In this lecture, we will get more practice with row reduction and in the process introduce two important types of matrix forms. We will also discuss when a linear system has a unique solution, infinitely many solutions, or no solution. Lastly, we will introduce a convenient parameter called the rank of a matrix.

2.1 Row echelon form (REF)

Consider the linear system

$$\begin{aligned} x_1 + 5x_2 - 2x_4 - x_5 + 7x_6 &= -4 \\ 2x_2 - 2x_3 + 3x_6 &= 0 \\ -9x_4 - x_5 + x_6 &= -1 \\ 5x_5 + x_6 &= 5 \\ 0 &= 0 \end{aligned}$$

having augmented matrix

$$\begin{bmatrix} 1 & 5 & 0 & -2 & -1 & 7 & -4 \\ 0 & 2 & -2 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -9 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 5 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The above augmented matrix has the following properties:

P1. All nonzero rows are above any rows of all zeros.

P2. The leftmost nonzero entry of a row is to the right of the leftmost nonzero entry of the row above it.

www.RanaMaths.com

Any matrix satisfying properties P1 and P2 is said to be in **row echelon form** (**REF**). In REF, the leftmost nonzero entry in a row is called a **leading entry**:

$$\begin{bmatrix} \mathbf{1} & 5 & 0 & -2 & -1 & 7 & -4 \\ 0 & \mathbf{2} & -2 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -\mathbf{9} & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & \mathbf{5} & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

A consequence of property P2 is that every entry below a leading entry is zero:

$$\begin{bmatrix} \mathbf{1} & 5 & 0 & -2 & -4 & -1 & -7 \\ \mathbf{0} & \mathbf{2} & -2 & 0 & 0 & 3 & 0 \\ \mathbf{0} & \mathbf{0} & 0 & -\mathbf{9} & -1 & 1 & -1 \\ \mathbf{0} & \mathbf{0} & 0 & \mathbf{0} & \mathbf{5} & 1 & 5 \\ \mathbf{0} & \mathbf{0} & 0 & \mathbf{0} & \mathbf{0} & 0 & 0 \end{bmatrix}$$

We can perform elementary row operations, or **row reduction**, to transform a matrix into REF.

Example 2.1. Explain why the following matrices are not in REF. Use elementary row operations to put them in REF.

$$\mathbf{M} = \begin{bmatrix} 3 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 \end{bmatrix} \quad \mathbf{N} = \begin{bmatrix} 7 & 5 & 0 & -3 \\ 0 & 3 & -1 & 1 \\ 0 & 6 & -5 & 2 \end{bmatrix}$$

Solution. Matrix $\mathbf{M}$ fails property P1. To put $\mathbf{M}$ in REF we interchange R_2 with R_3:

$$\mathbf{M} = \begin{bmatrix} 3 & -1 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 \end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 3 & -1 & 0 & 3 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The matrix $\mathbf{N}$ fails property P2. To put $\mathbf{N}$ in REF we perform the operation $-2R_2 + R_3 \rightarrow R_3$:

$$\begin{bmatrix} 7 & 5 & 0 & -3 \\ 0 & 3 & -1 & 1 \\ 0 & 6 & -5 & 2 \end{bmatrix} \xrightarrow{-2R_2 + R_3} \begin{bmatrix} 7 & 5 & 0 & -3 \\ 0 & 3 & -1 & 1 \\ 0 & 0 & -3 & 0 \end{bmatrix}$$

□

Why is REF useful? Certain properties of a matrix can be easily deduced if it is in REF. For now, REF is useful to us for solving a linear system of equations. If an augmented matrix is in REF, we can use **back substitution** to solve the system, just as we did in Lecture 1. For example, consider the system

$$\begin{aligned} 8x_1 - 2x_2 + x_3 &= 4 \\ 3x_2 - x_3 &= 7 \\ 2x_3 &= 4 \end{aligned}$$

www.RanaMaths.com

whose augmented matrix is already in REF:

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

From the last equation we obtain that $2x_3 = 4$, and thus $x_3 = 2$. Substituting $x_3 = 2$ into the second equation we obtain that $x_2 = 3$. Substituting $x_3 = 2$ and $x_2 = 3$ into the first equation we obtain that $x_1 = 1$.

2.2 Reduced row echelon form (RREF)

Although REF simplifies the problem of solving a linear system, later on in the course we will need to completely row reduce matrices into what is called **reduced row echelon form (RREF)**. A matrix is in RREF if it is in REF (so it satisfies properties P1 and P2) and in addition satisfies the following properties:

P3. The leading entry in each nonzero row is a 1.

P4. All the entries above (and below) a leading 1 are all zero.

A leading 1 in the RREF of a matrix is called a **pivot**. For example, the following matrix in RREF:

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

has three pivots:

$$\begin{bmatrix} \mathbf{1} & 6 & \mathbf{0} & 3 & \mathbf{0} & 0 \\ \mathbf{0} & 0 & \mathbf{1} & -4 & \mathbf{0} & 5 \\ \mathbf{0} & 0 & \mathbf{0} & 0 & \mathbf{1} & 7 \end{bmatrix}$$

Example 2.2. Use row reduction to transform the matrix into RREF.

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

Solution. The first step is to make the top leftmost entry nonzero:

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix} \xrightarrow{R_3 \leftrightarrow R_1} \begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Now create a leading 1 in the first row:

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \xrightarrow{\frac{1}{3}R_1} \begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

www.RanaMaths.com

Create zeros under the newly created leading 1:

$$\begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \xrightarrow{-3R_1+R_2} \begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Create a leading 1 in the second row:

$$\begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \xrightarrow{\frac{1}{2}R_2} \begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 0 & 1 & -2 & 2 & 1 & -3 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Create zeros under the newly created leading 1:

$$\begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 0 & 1 & -2 & 2 & 1 & -3 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix} \xrightarrow{-3R_2+R_3} \begin{bmatrix} 1 & -3 & 4 & -3 & 2 & 5 \\ 0 & 1 & -2 & 2 & 1 & -3 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

We have now completed the top-to-bottom phase of the row reduction algorithm. In the next phase, we work bottom-to-top and create zeros **above** the leading 1's. Create zeros above the leading 1 in the third row:

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

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

Create zeros above the leading 1 in the second row:

$$\begin{bmatrix} 1 & -3 & 4 & -3 & 0 & -3 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \xrightarrow{3R_2+R_1} \begin{bmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

This completes the row reduction algorithm and the matrix is in RREF. □

Example 2.3. Use row reduction to solve the linear system.

$$\begin{aligned} 2x_1 + 4x_2 + 6x_3 &= 8 \\ x_1 + 2x_2 + 4x_3 &= 8 \\ 3x_1 + 6x_2 + 9x_3 &= 12 \end{aligned}$$

Solution. The augmented matrix is

$$\begin{bmatrix} 2 & 4 & 6 & 8 \\ 1 & 2 & 4 & 8 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

www.RanaMaths.com

Create a leading 1 in the first row:

$$\begin{bmatrix} 2 & 4 & 6 & 8 \\ 1 & 2 & 4 & 8 \\ 3 & 6 & 9 & 12 \end{bmatrix} \xrightarrow{\frac{1}{2}R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 8 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

Create zeros under the first leading 1:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 8 \\ 3 & 6 & 9 & 12 \end{bmatrix} \xrightarrow{-R_1+R_2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 4 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 4 \\ 3 & 6 & 9 & 12 \end{bmatrix} \xrightarrow{-3R_1+R_3} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The system is consistent, however, there are only 2 nonzero rows but 3 unknown variables. This means that the solution set will contain $3 - 2 = 1$ **free parameter**. The second row in the augmented matrix is equivalent to the equation:

$$x_3 = 4.$$

The first row is equivalent to the equation:

$$x_1 + 2x_2 + 3x_3 = 4$$

and after substituting $x_3 = 4$ we obtain

$$x_1 + 2x_2 = -8.$$

We now must choose one of the variables x_1 or x_2 to be a parameter, say t, and solve for the remaining variable. If we set $x_2 = t$ then from $x_1 + 2x_2 = -8$ we obtain that

$$x_1 = -8 - 2t.$$

We can therefore write the solution set for the linear system as

$$\begin{aligned} x_1 &= -8 - 2t \\ x_2 &= t \\ x_3 &= 4 \end{aligned} \qquad (2.1)$$

where t can be any real number. If we had chosen x_1 to be the parameter, say $x_1 = t$, then the solution set can be written as

$$\begin{aligned} x_1 &= t \\ x_2 &= -4 - \tfrac{1}{2}t \\ x_3 &= 4 \end{aligned} \qquad (2.2)$$

Although (2.1) and (2.2) are two different parameterizations, they both give the same solution set. □

In general, if a linear system has n unknown variables and the row reduced augmented matrix has r leading entries, then the number of free parameters d in the solution set is

$$d = n - r.$$

Thus, when performing back substitution, we will have to set d of the unknown variables to arbitrary parameters. In the previous example, there are $n = 3$ unknown variables and the row reduced augmented matrix contained $r = 2$ leading entries. The number of free parameters was therefore

$$d = n - r = 3 - 2 = 1.$$

Because the number of leading entries r in the row reduced coefficient matrix determine the number of free parameters, we will refer to r as the **rank** of the coefficient matrix:

$$r = \text{rank}(\mathbf{A}).$$

Later in the course, we will give a more geometric interpretation to $\text{rank}(\mathbf{A})$.

Example 2.4. Solve the linear system represented by the augmented matrix

$$\begin{bmatrix} 1 & -7 & 2 & -5 & 8 & 10 \\ 0 & 1 & -3 & 3 & 1 & -5 \\ 0 & 0 & 0 & 1 & -1 & 4 \end{bmatrix}$$

Solution. The number of unknowns is $n = 5$ and the augmented matrix has rank $r = 3$ (leading entries). Thus, the solution set is parameterized by $d = 5 - 3 = 2$ free variables, call them t and s. The last equation of the augmented matrix is $x_4 - x_5 = 4$. We choose x_5 to be the first parameter so we set $x_5 = t$. Therefore, $x_4 = 4 + t$. The second equation of the augmented matrix is

$$x_2 - 3x_3 + 3x_4 + x_5 = -5$$

and the unassigned variables are x_2 and x_3. We choose x_3 to be the second parameter, say $x_3 = s$. Then

$$\begin{aligned} x_2 &= -5 + 3x_3 - 3x_4 - x_5 \\ &= -5 + 3s - 3(4 + t) - t \\ &= -17 - 4t + 3s. \end{aligned}$$

We now use the first equation of the augmented matrix to write x_1 in terms of the other variables:

$$\begin{aligned} x_1 &= 10 + 7x_2 - 2x_3 + 5x_4 - 8x_5 \\ &= 10 + 7(-17 - 4t + 3s) - 2s + 5(4 + t) - 8t \\ &= -89 - 31t + 19s \end{aligned}$$

www.RanaMaths.com

Thus, the solution set is

$$\begin{aligned} x_1 &= -89 - 31t + 19s \\ x_2 &= -17 - 4t + 3s \\ x_3 &= s \\ x_4 &= 4 + t \\ x_5 &= t \end{aligned}$$

where t and s are arbitrary real numbers.. Choose arbitrary numbers for t and s and substitute the corresponding list $(x_1, x_2, \ldots, x_5)$ into the system of equations to verify that it is a solution. □

2.3 Existence and uniqueness of solutions

The REF or RREF of an augmented matrix leads to three distinct possibilities for the solution set of a linear system.

Theorem 2.5: Let $[\mathbf{A}\ \mathbf{b}]$ be the augmented matrix of a linear system. One of the following distinct possibilities will occur:

1. The augmented matrix will contain an inconsistent row.
2. All the rows of the augmented matrix are consistent and there are no free parameters.
3. All the rows of the augmented matrix are consistent and there are $d \geq 1$ variables that must be set to arbitrary parameters

In Case 1., the linear system is inconsistent and thus has no solution. In Case 2., the linear system is consistent and has only one (and thus **unique**) solution. This case occurs when $r = \text{rank}(\mathbf{A}) = n$ since then the number of free parameters is $d = n - r = 0$. In Case 3., the linear system is consistent and has infinitely many solutions. This case occurs when $r < n$ and thus $d = n - r > 0$ is the number of free parameters.

After this lecture you should know the following:

- what the REF is and how to compute it
- what the RREF is and how to compute it
- how to solve linear systems using row reduction (**Practice!!!**)
- how to identify when a linear system is inconsistent
- how to identify when a linear system is consistent
- what is the rank of a matrix
- how to compute the number of free parameters in a solution set
- what are the three possible cases for the solution set of a linear system (Theorem 2.5)

www.RanaMaths.com

www.RanaMaths.com

Lecture 3

Vector Equations

In this lecture, we introduce vectors and vector equations. Specifically, we introduce the linear combination problem which simply asks whether it is possible to express one vector in terms of other vectors; we will be more precise in what follows. As we will see, solving the linear combination problem reduces to solving a linear system of equations.

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

Recall that a **column vector** in $\mathbb{R}^n$ is a $n \times 1$ matrix. From now on, we will drop the "column" descriptor and simply use the word **vectors**. It is important to emphasize that a vector in $\mathbb{R}^n$ is simply a list of n numbers; you are safe (and highly encouraged!) to forget the idea that a vector is an object with an arrow. Here is a vector in $\mathbb{R}^2$:

$$\mathbf{v} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

Here is a vector in $\mathbb{R}^3$:

$$\mathbf{v} = \begin{bmatrix} -3 \\ 0 \\ 11 \end{bmatrix}.$$

Here is a vector in $\mathbb{R}^6$:

$$\mathbf{v} = \begin{bmatrix} 9 \\ 0 \\ -3 \\ 6 \\ 0 \\ 3 \end{bmatrix}.$$

To indicate that $\mathbf{v}$ is a vector in $\mathbb{R}^n$, we will use the notation $\mathbf{v} \in \mathbb{R}^n$. The mathematical symbol $\in$ means "is an element of". When we write vectors within a paragraph, we will write them using list notation instead of column notation, e.g., $\mathbf{v} = (-1, 4)$ instead of $\mathbf{v} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$.

www.RanaMaths.com

We can add/subtract vectors, and multiply vectors by numbers or **scalars**. For example, here is the addition of two vectors:

$$\begin{bmatrix} 0 \\ -5 \\ 9 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ -3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -8 \\ 9 \\ 3 \end{bmatrix}.$$

And the multiplication of a scalar with a vector:

$$3 \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \\ 15 \end{bmatrix}.$$

And here are both operations combined:

$$-2 \begin{bmatrix} 4 \\ -8 \\ 3 \end{bmatrix} + 3 \begin{bmatrix} -2 \\ 9 \\ 4 \end{bmatrix} = \begin{bmatrix} -8 \\ 16 \\ -6 \end{bmatrix} + \begin{bmatrix} -6 \\ 27 \\ 12 \end{bmatrix} = \begin{bmatrix} -14 \\ 43 \\ 6 \end{bmatrix}.$$

These operations constitute "the algebra" of vectors. As the following example illustrates, vectors can be used in a natural way to represent the solution of a linear system.

Example 3.1. Write the general solution in vector form of the linear system represented by the augmented matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & -7 & 2 & -5 & 8 & 10 \\ 0 & 1 & -3 & 3 & 1 & -5 \\ 0 & 0 & 0 & 1 & -1 & 4 \end{bmatrix}$$

Solution. The number of unknowns is $n = 5$ and the associated coefficient matrix $\mathbf{A}$ has rank $r = 3$. Thus, the solution set is parametrized by $d = n - r = 2$ parameters. This system was considered in Example 2.4 and the general solution was found to be

$$\begin{aligned} x_1 &= -89 - 31t_1 + 19t_2 \\ x_2 &= -17 - 4t_1 + 3t_2 \\ x_3 &= t_2 \\ x_4 &= 4 + t_1 \\ x_5 &= t_1 \end{aligned}$$

where t_1 and t_2 are arbitrary real numbers. The solution in vector form therefore takes the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -89 - 31t_1 + 19t_2 \\ -17 - 4t_1 + 3t_2 \\ t_2 \\ 4 + t_1 \\ t_1 \end{bmatrix} = \begin{bmatrix} -89 \\ -17 \\ 0 \\ 4 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} -31 \\ -4 \\ 0 \\ 1 \\ 1 \end{bmatrix} + t_2 \begin{bmatrix} 19 \\ 3 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

□

www.RanaMaths.com

A fundamental problem in **linear algebra** is solving vector equations for an unknown vector. As an example, suppose that you are given the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 4 \\ -8 \\ 3 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} -2 \\ 9 \\ 4 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} -14 \\ 43 \\ 6 \end{bmatrix},$$

and asked to find numbers x_1 and x_2 such that $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \mathbf{b}$, that is,

$$x_1 \begin{bmatrix} 4 \\ -8 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 9 \\ 4 \end{bmatrix} = \begin{bmatrix} -14 \\ 43 \\ 6 \end{bmatrix}.$$

Here the unknowns are the scalars x_1 and x_2. After some guess and check, we find that $x_1 = -2$ and $x_2 = 3$ is a solution to the problem since

$$-2 \begin{bmatrix} 4 \\ -8 \\ 3 \end{bmatrix} + 3 \begin{bmatrix} -2 \\ 9 \\ 4 \end{bmatrix} = \begin{bmatrix} -14 \\ 43 \\ 6 \end{bmatrix}.$$

In some sense, the vector $\mathbf{b}$ is a combination of the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. This motivates the following definition.

Definition 3.2: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors in $\mathbb{R}^n$. A vector $\mathbf{b}$ is said to be a **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ if there exists scalars $x_1, x_2, \ldots, x_p$ such that $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{b}$.

The scalars in a linear combination are called the **coefficients** of the linear combination. As an example, given the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} -2 \\ 4 \\ -6 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} -1 \\ 5 \\ 6 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} -3 \\ 0 \\ -27 \end{bmatrix}$$

you can verify (and you should!) that

$$3\mathbf{v}_1 + 4\mathbf{v}_2 - 2\mathbf{v}_3 = \mathbf{b}.$$

Therefore, we can say that $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ with coefficients $x_1 = 3$, $x_2 = 4$, and $x_3 = -2$.

3.2 The linear combination problem

The linear combination problem is the following:

www.RanaMaths.com

Problem: Given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ and $\mathbf{b}$, is $\mathbf{b}$ a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$?

For example, say you are given the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

and also

$$\mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}.$$

Does there exist scalars x_1, x_2, x_3 such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}? \tag{3.1}$$

For obvious reasons, equation (3.1) is called a **vector equation** and the unknowns are x_1, x_2, and x_3. To gain some intuition with the linear combination problem, let's do an example by inspection.

Example 3.3. Let $\mathbf{v}_1 = (1, 0, 0)$, let $\mathbf{v}_2 = (0, 0, 1)$, let $\mathbf{b}_1 = (0, 2, 0)$, and let $\mathbf{b}_2 = (-3, 0, 7)$. Are $\mathbf{b}_1$ and $\mathbf{b}_2$ linear combinations of $\mathbf{v}_1, \mathbf{v}_2$?

Solution. For any scalars x_1 and x_2

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ x_2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

and thus no, $\mathbf{b}_1$ is not a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. On the other hand, by inspection we have that

$$-3\mathbf{v}_1 + 7\mathbf{v}_2 = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 7 \end{bmatrix} = \mathbf{b}_2$$

and thus yes, $\mathbf{b}_2$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. These examples, of low dimension, were more-or-less obvious. Going forward, we are going to need a systematic way to solve the linear combination problem that does not rely on pure inspection. □

We now describe how the linear combination problem is connected to the problem of solving a system of linear equations. Consider again the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}.$$

Does there exist scalars x_1, x_2, x_3 such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}? \tag{3.2}$$

www.RanaMaths.com

First, let's expand the left-hand side of equation (3.2):

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \begin{bmatrix} x_1 \\ 2x_1 \\ x_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 2x_3 \\ x_3 \\ 2x_3 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 + 2x_3 \\ 2x_1 + x_2 + x_3 \\ x_1 + 2x_3 \end{bmatrix}.$$

We want equation (3.2) to hold so let's equate the expansion $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3$ with $\mathbf{b}$. In other words, set

$$\begin{bmatrix} x_1 + x_2 + 2x_3 \\ 2x_1 + x_2 + x_3 \\ x_1 + 2x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}.$$

Comparing component-by-component in the above relationship, we seek scalars x_1, x_2, x_3 satisfying the equations

$$\begin{aligned} x_1 + x_2 + 2x_3 &= 0 \\ 2x_1 + x_2 + x_3 &= 1 \\ x_1 + 2x_3 &= -2. \end{aligned} \tag{3.3}$$

This is just a linear system consisting of $m = 3$ equations and $n = 3$ unknowns! Thus, the linear combination problem can be solved by solving a system of linear equations for the unknown scalars x_1, x_2, x_3. We know how to do this. In this case, the augmented matrix of the linear system (3.3) is

$$[\mathbf{A}\ \mathbf{b}] = \begin{bmatrix} 1 & 1 & 2 & 0 \\ 2 & 1 & 1 & 1 \\ 1 & 0 & 2 & -2 \end{bmatrix}$$

Notice that the 1st column of $\mathbf{A}$ is just $\mathbf{v}_1$, the second column is $\mathbf{v}_2$, and the third column is $\mathbf{v}_3$, in other words, the augment matrix is

$$[\mathbf{A}\ \mathbf{b}] = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{b} \end{bmatrix}$$

Applying the row reduction algorithm, the solution is

$$x_1 = 0,\ x_2 = 2,\ x_3 = -1$$

and thus these coefficients solve the linear combination problem. In other words,

$$0\mathbf{v}_1 + 2\mathbf{v}_2 - \mathbf{v}_3 = \mathbf{b}$$

In this case, there is only one solution to the linear system, so $\mathbf{b}$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in only one (or unique) way. You should verify these computations.

We summarize the previous discussion with the following:

The problem of determining if a given vector $\mathbf{b}$ is a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ is equivalent to solving the linear system of equations with augmented matrix

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p & \mathbf{b} \end{bmatrix}.$$

www.RanaMaths.com

Applying the existence and uniqueness Theorem 2.5, the only three possibilities to the linear combination problem are:

1. If the linear system is inconsistent then $\mathbf{b}$ is not a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, i.e., there does not exist scalars $x_1, x_2, \ldots, x_p$ such that $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{b}$.

2. If the linear system is consistent and the solution is unique then $\mathbf{b}$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in only one way.

3. If the the linear system is consistent and the solution set has free parameters, then $\mathbf{b}$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in infinitely many ways.

Example 3.4. Is the vector $\mathbf{b} = (7, 4, -3)$ a linear combination of the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}?$$

Solution. Form the augmented matrix:

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix}$$

The RREF of the augmented matrix is

$$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

and therefore the solution is $x_1 = 3$ and $x_2 = 2$. Therefore, yes, $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2$:

$$3\mathbf{v}_1 + 2\mathbf{v}_2 = 3\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix} = \mathbf{b}$$

Notice that the solution set does not contain any free parameters because $n = 2$ (unknowns) and $r = 2$ (rank) and so $d = 0$. Therefore, the above linear combination is the only way to write $\mathbf{b}$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. □

Example 3.5. Is the vector $\mathbf{b} = (1, 0, 1)$ a linear combination of the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}?$$

www.RanaMaths.com

Solution. The augmented matrix of the corresponding linear system is

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 2 & 0 & 4 & 1 \end{bmatrix}.$$

After row reducing we obtain that

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

The last row is inconsistent, and therefore the linear system does not have a solution. Therefore, no, $\mathbf{b}$ is not a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. □

Example 3.6. Is the vector $\mathbf{b} = (8, 8, 12)$ a linear combination of the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 4 \\ 9 \end{bmatrix}?$$

Solution. The augmented matrix is

$$\begin{bmatrix} 2 & 4 & 6 & 8 \\ 1 & 2 & 4 & 8 \\ 3 & 6 & 9 & 12 \end{bmatrix} \xrightarrow{\text{REF}} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The system is consistent and therefore $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. In this case, the solution set contains $d = 1$ free parameters and therefore, it is possible to write $\mathbf{b}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in infinitely many ways. In terms of the parameter t, the solution set is

$$\begin{aligned} x_1 &= -8 - 2t \\ x_2 &= t \\ x_3 &= 4 \end{aligned}$$

Choosing any t gives scalars that can be used to write $\mathbf{b}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. For example, choosing $t = 1$ we obtain $x_1 = -10$, $x_2 = 1$, and $x_3 = 4$, and you can verify that

$$-10\mathbf{v}_1 + \mathbf{v}_2 + 4\mathbf{v}_3 = -10 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix} + 4 \begin{bmatrix} 6 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 12 \end{bmatrix} = \mathbf{b}$$

Or, choosing $t = -2$ we obtain $x_1 = -4$, $x_2 = -2$, and $x_3 = 4$, and you can verify that

$$-4\mathbf{v}_1 - 2\mathbf{v}_2 + 4\mathbf{v}_3 = -4 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - 2 \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix} + 4 \begin{bmatrix} 6 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 12 \end{bmatrix} = \mathbf{b}$$

□

www.RanaMaths.com

We make a few important observations on linear combinations of vectors. Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, there are certain vectors $\mathbf{b}$ that can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in an obvious way. The zero vector $\mathbf{b} = \mathbf{0}$ can always be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$:

$$\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p.$$

Each $\mathbf{v}_i$ itself can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, for example,

$$\mathbf{v}_2 = 0\mathbf{v}_1 + (1)\mathbf{v}_2 + 0\mathbf{v}_3 + \cdots + 0\mathbf{v}_p.$$

More generally, any scalar multiple of $\mathbf{v}_i$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, for example,

$$x\mathbf{v}_2 = 0\mathbf{v}_1 + x\mathbf{v}_2 + 0\mathbf{v}_3 + \cdots + 0\mathbf{v}_p.$$

By varying the coefficients $x_1, x_2, \ldots, x_p$, we see that there are infinitely many vectors $\mathbf{b}$ that can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. The "space" of all the possible linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ has a name, which we introduce next.

3.3 The span of a set of vectors

Given a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, we have been considering the problem of whether or not a given vector $\mathbf{b}$ is a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$. We now take another point of view and instead consider the idea of **generating** all vectors that are a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$. So how do we generate a vector that is guaranteed to be a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$? For example, if $\mathbf{v}_1 = (2, 1, 3)$, $\mathbf{v}_2 = (4, 2, 6)$ and $\mathbf{v}_3 = (6, 4, 9)$ then

$$-10\mathbf{v}_1 + \mathbf{v}_2 + 4\mathbf{v}_3 = -10 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix} + 4 \begin{bmatrix} 6 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 12 \end{bmatrix}.$$

Thus, by construction, the vector $\mathbf{b} = (8, 8, 12)$ is a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. This discussion leads us to the following definition.

Definition 3.7: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors. The set of all vectors that are a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ is called the **span** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, and we denote it by

$$S = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}.$$

By definition, the span of a set of vectors is a collection of vectors, or a **set** of vectors. If $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ then $\mathbf{b}$ is an **element** of the set $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, and we write this as

$$\mathbf{b} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}.$$

www.RanaMaths.com

By definition, writing that $\mathbf{b} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ implies that there exists scalars $x_1, x_2, \ldots, x_p$ such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{b}.$$

Even though $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is an infinite set of vectors, it is not necessarily true that it is the whole space $\mathbb{R}^n$.

The set $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is just a collection of infinitely many vectors but it has some geometric structure. In $\mathbb{R}^2$ and $\mathbb{R}^3$ we can visualize $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$. In $\mathbb{R}^2$, the span of a single nonzero vector, say $\mathbf{v} \in \mathbb{R}^2$, is a line through the origin in the direction of $\mathbf{v}$, see Figure 3.1.

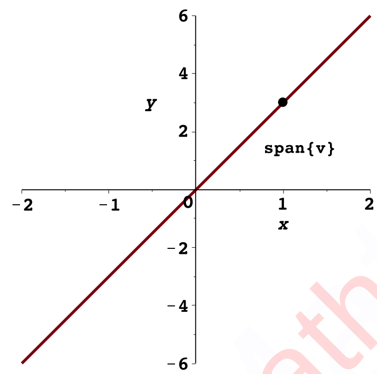


Figure 3.1: The span of a single non-zero vector in $\mathbb{R}^2$.

In $\mathbb{R}^2$, the span of two vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ that are not multiples of each other is all of $\mathbb{R}^2$. That is, $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \mathbb{R}^2$. For example, with $\mathbf{v}_1 = (1, 0)$ and $\mathbf{v}_2 = (0, 1)$, it is true that $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \mathbb{R}^2$. In $\mathbb{R}^3$, the span of two vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^3$ that are not multiples of each other is a plane through the origin containing $\mathbf{v}_1$ and $\mathbf{v}_2$, see Figure 3.2. In $\mathbb{R}^3$, the

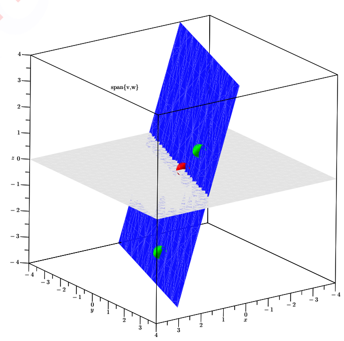


Figure 3.2: The span of two vectors, not multiples of each other, in $\mathbb{R}^3$.

span of a single vector is a line through the origin, and the span of three vectors that do not depend on each other (we will make this precise soon) is all of $\mathbb{R}^3$.

Example 3.8. Is the vector $\mathbf{b} = (7, 4, -3)$ in the span of the vectors $\mathbf{v}_1 = (1, -2, -5), \mathbf{v}_2 = (2, 5, 6)$? In other words, is $\mathbf{b} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$?

www.RanaMaths.com

Solution. By definition, $\mathbf{b}$ is in the span of $\mathbf{v}_1$ and $\mathbf{v}_2$ if there exists scalars x_1 and x_2 such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \mathbf{b},$$

that is, if $\mathbf{b}$ can be written as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. From our previous discussion on the linear combination problem, we must consider the augmented matrix $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{b}\end{bmatrix}$. Using row reduction, the augmented matrix is consistent and there is only one solution (see Example 3.4). Therefore, yes, $\mathbf{b} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and the linear combination is unique. □

Example 3.9. Is the vector $\mathbf{b} = (1, 0, 1)$ in the span of the vectors $\mathbf{v}_1 = (1, 0, 2), \mathbf{v}_2 = (0, 1, 0), \mathbf{v}_3 = (2, 1, 4)$?

Solution. From Example 3.5, we have that

$$\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{b}\end{bmatrix} \xrightarrow{\text{REF}} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

The last row is inconsistent and therefore $\mathbf{b}$ is not in $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. □

Example 3.10. Is the vector $\mathbf{b} = (8, 8, 12)$ in the span of the vectors $\mathbf{v}_1 = (2, 1, 3), \mathbf{v}_2 = (4, 2, 6), \mathbf{v}_3 = (6, 4, 9)$?

Solution. From Example 3.6, we have that

$$\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{b}\end{bmatrix} \xrightarrow{\text{REF}} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The system is consistent and therefore $\mathbf{b} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. In this case, the solution set contains $d = 1$ free parameters and therefore, it is possible to write $\mathbf{b}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in infinitely many ways. □

Example 3.11. Answer the following with True or False, and explain your answer.

(a) The vector $\mathbf{b} = (1, 2, 3)$ is in the span of the set of vectors

$$\left\{ \begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -7 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ -5 \\ 0 \end{bmatrix} \right\}.$$

(b) The solution set of the linear system whose augmented matrix is $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{b}\end{bmatrix}$ is the same as the solution set of the vector equation $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}$.

(c) Suppose that the augmented matrix $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{b}\end{bmatrix}$ has an inconsistent row. Then either $\mathbf{b}$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ or $\mathbf{b} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

(d) The span of the vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ (at least one of which is nonzero) contains only the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ and the zero vector $\mathbf{0}$.

www.RanaMaths.com

After this lecture you should know the following:

- what a vector is
- what a linear combination of vectors is
- what the linear combination problem is
- the relationship between the linear combination problem and the problem of solving linear systems of equations
- how to solve the linear combination problem
- what the span of a set of vectors is
- the relationship between what it means for a vector $\mathbf{b}$ to be in the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ and the problem of writing $\mathbf{b}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$
- the geometric interpretation of the span of a set of vectors

www.RanaMaths.com

www.RanaMaths.com

Lecture 4

The Matrix Equation $\mathbf{Ax} = \mathbf{b}$

In this lecture, we introduce the operation of matrix-vector multiplication and how it relates to the linear combination problem.

4.1 Matrix-vector multiplication

We begin with the definition of matrix-vector multiplication.

Definition 4.1: Given a matrix $\mathbf{A} \in M_{m\times n}$ and a vector $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

we define the product of $\mathbf{A}$ and $\mathbf{x}$ as the vector $\mathbf{Ax}$ in $\mathbb{R}^m$ given by

$$\mathbf{Ax} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x}} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

For the product $\mathbf{Ax}$ to be well-defined, the number of columns of $\mathbf{A}$ must equal the number of components of $\mathbf{x}$. Another way of saying this is that the outer dimension of $\mathbf{A}$ must equal the inner dimension of $\mathbf{x}$:

$$(m \times n) \cdot (n \times 1) \to m \times 1$$

Example 4.2. Compute $\mathbf{Ax}$.

www.RanaMaths.com

(a)

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 3 & 0 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 2 \\ -4 \\ -3 \\ 8 \end{bmatrix}$$

(b)

$$\mathbf{A} = \begin{bmatrix} 3 & 3 & -2 \\ 4 & -4 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

(c)

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 4 & 1 & -2 \\ 3 & -3 & 3 \\ 0 & -2 & -3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} -1 \\ 2 \\ -2 \end{bmatrix}$$

Solution. We compute:

(a)

$$\begin{aligned} \mathbf{Ax} &= \begin{bmatrix} 1 & -1 & 3 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ -4 \\ -3 \\ 8 \end{bmatrix} \\ &= \begin{bmatrix} (1)(2) + (-1)(-4) + (3)(-3) + (0)(8) \end{bmatrix} = \begin{bmatrix} -3 \end{bmatrix} \end{aligned}$$

(b)

$$\begin{aligned} \mathbf{Ax} &= \begin{bmatrix} 3 & 3 & -2 \\ 4 & -4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} (3)(1) + (3)(0) + (-2)(-1) \\ (4)(1) + (-4)(0) + (-1)(-1) \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 5 \end{bmatrix} \end{aligned}$$

(c)

$$\begin{aligned}
\mathbf{Ax} &= \begin{bmatrix} -1 & 1 & 0 \\ 4 & 1 & -2 \\ 3 & -3 & 3 \\ 0 & -2 & -3 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -2 \end{bmatrix} \\
&= \begin{bmatrix} (-1)(-1) + (1)(2) + (0)(-2) \\ (4)(-1) + (1)(2) + (-2)(-2) \\ (3)(-1) + (-3)(2) + (3)(-2) \\ (0)(-1) + (-2)(2) + (-3)(-2) \end{bmatrix} \\
&= \begin{bmatrix} 3 \\ 2 \\ -15 \\ 2 \end{bmatrix}
\end{aligned}$$

□

We now list two important properties of matrix-vector multiplication.

Theorem 4.3: Let $\mathbf{A}$ be an $m \times n$ a matrix.

(a) For any vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ it holds that

$$\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av}.$$

(b) For any vector $\mathbf{u}$ and scalar c it holds that

$$\mathbf{A}(c\mathbf{u}) = c(\mathbf{Au}).$$

Example 4.4. For the given data, verify that the properties of Theorem 4.3 hold:

$$\mathbf{A} = \begin{bmatrix} 3 & -3 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad c = -2.$$

4.2 Matrix-vector multiplication and linear combinations

Recall the general definition of matrix-vector multiplication $\mathbf{Ax}$ is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} \tag{4.1}$$

There is an important way to decompose matrix-vector multiplication involving a linear combination. To see how, let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ denote the columns of $\mathbf{A}$ and consider the following linear combination:

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \begin{bmatrix} x_1a_{11} \\ x_1a_{21} \\ \vdots \\ x_1a_{m1} \end{bmatrix} + \begin{bmatrix} x_2a_{12} \\ x_2a_{22} \\ \vdots \\ x_2a_{m2} \end{bmatrix} + \cdots + \begin{bmatrix} x_na_{1n} \\ x_na_{2n} \\ \vdots \\ x_na_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} x_1a_{11} + x_2a_{12} + \cdots + x_na_{1n} \\ x_1a_{21} + x_2a_{22} + \cdots + x_na_{2n} \\ \vdots \\ x_1a_{m1} + x_2a_{m2} + \cdots + x_na_{mn} \end{bmatrix}. \tag{4.2}$$

We observe that expressions (4.1) and (4.2) are equal! Therefore, if $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ then

$$\mathbf{Ax} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n.$$

In summary, the vector $\mathbf{Ax}$ is a linear combination of the columns of $\mathbf{A}$ where the scalar in the linear combination are the components of $\mathbf{x}$! This (important) observation gives an alternative way to compute $\mathbf{Ax}$.

Example 4.5. Given

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 4 & 1 & -2 \\ 3 & -3 & 3 \\ 0 & -2 & -3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} -1 \\ 2 \\ -2 \end{bmatrix},$$

compute $\mathbf{Ax}$ in two ways: (1) using the original Definition 4.1, and (2) as a linear combination of the columns of $\mathbf{A}$.

4.3 The matrix equation problem

As we have seen, with a matrix $\mathbf{A}$ and any vector $\mathbf{x}$, we can produce a new output vector via the multiplication $\mathbf{Ax}$. If $\mathbf{A}$ is a $m \times n$ matrix then we must have $\mathbf{x} \in \mathbb{R}^n$ and the output vector $\mathbf{Ax}$ is in $\mathbb{R}^m$. We now introduce the following problem:

Problem: Given a matrix $\mathbf{A} \in M_{m\times n}$ and a vector $\mathbf{b} \in \mathbb{R}^m$, find, if possible, a vector $\mathbf{x} \in \mathbb{R}^n$ such that

$$\mathbf{Ax} = \mathbf{b}. \tag{$\star$}$$

Equation ($\star$) is a **matrix equation** where the unknown variable is $\mathbf{x}$. If $\mathbf{u}$ is a vector such that $\mathbf{Au} = \mathbf{b}$, then we say that $\mathbf{u}$ is a solution to the equation $\mathbf{Ax} = \mathbf{b}$. For example,

www.RanaMaths.com

suppose that

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \; \mathbf{b} = \begin{bmatrix} -3 \\ 7 \end{bmatrix}.$$

Does the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ have a solution? Well, for any $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ we have that

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

and thus any output vector $\mathbf{A}\mathbf{x}$ has equal entries. Since $\mathbf{b}$ does not have equal entries then the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has no solution.

We now describe a systematic way to solve matrix equations. As we have seen, the vector $\mathbf{A}\mathbf{x}$ is a linear combination of the columns of $\mathbf{A}$ with the coefficients given by the components of $\mathbf{x}$. Therefore, the matrix equation problem is equivalent to the linear combination problem. In Lecture 2, we showed that the linear combination problem can be solved by solving a system of linear equations. Putting all this together then, if $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $\mathbf{b} \in \mathbb{R}^m$ then:

> To find a vector $\mathbf{x} \in \mathbb{R}^n$ that solves the matrix equation
>
> $$\mathbf{A}\mathbf{x} = \mathbf{b}$$
>
> we solve the linear system whose augmented matrix is
>
> $$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n & \mathbf{b} \end{bmatrix}.$$

From now on, a system of linear equations such as

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\ & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

will be written in the compact form

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A}$ is the coefficient matrix of the linear system, $\mathbf{b}$ is the output vector, and $\mathbf{x}$ is the unknown vector to be solved for. We summarize our findings with the following theorem.

> **Theorem 4.6:** Let $\mathbf{A} \in M_{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. The following statements are equivalent:
> (a) The equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution.
> (b) The vector $\mathbf{b}$ is a linear combination of the columns of $\mathbf{A}$.
> (c) The linear system represented by the augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ is consistent.

www.RanaMaths.com

Example 4.7. Solve, if possible, the matrix equation $\mathbf{Ax} = \mathbf{b}$ if

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -4 \\ 1 & 5 & 2 \\ -3 & -7 & -6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -2 \\ 4 \\ 12 \end{bmatrix}.$$

Solution. First form the augmented matrix:

$$[\mathbf{A}\ \mathbf{b}] = \begin{bmatrix} 1 & 3 & -4 & -2 \\ 1 & 5 & 2 & 4 \\ -3 & -7 & -6 & 12 \end{bmatrix}$$

Performing the row reduction algorithm we obtain that

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

Here $r = \text{rank}(\mathbf{A}) = 3$ and therefore $d = 0$, i.e., no free parameters. Peforming back substitution we obtain that $x_1 = -11$, $x_2 = 3$, and $x_3 = 0$. Thus, the solution to the matrix equation is unique (no free parameters) and is given by

$$\mathbf{x} = \begin{bmatrix} -11 \\ 3 \\ 0 \end{bmatrix}$$

Let's verify that $\mathbf{Ax} = \mathbf{b}$:

$$\mathbf{Ax} = \begin{bmatrix} 1 & 3 & -4 \\ 1 & 5 & 2 \\ -3 & -7 & -6 \end{bmatrix} \begin{bmatrix} -11 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -11 + 9 + 0 \\ -11 + 15 + 0 \\ 33 - 21 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ 12 \end{bmatrix} = \mathbf{b}$$

In other words, $\mathbf{b}$ is a linear combination of the columns of $\mathbf{A}$:

$$-11 \begin{bmatrix} 1 \\ 1 \\ -3 \end{bmatrix} + 3 \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix} + 0 \begin{bmatrix} -4 \\ 2 \\ -6 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ 12 \end{bmatrix}$$

□

www.RanaMaths.com

Example 4.8. Solve, if possible, the matrix equation $\mathbf{Ax} = \mathbf{b}$ if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}.$$

Solution. Row reducing the augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ we get

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -4 \end{bmatrix} \xrightarrow{-2R_1+R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & -10 \end{bmatrix}.$$

The last row is inconsistent and therefore there is no solution to the matrix equation $\mathbf{Ax} = \mathbf{b}$. In other words, $\mathbf{b}$ is not a linear combination of the columns of $\mathbf{A}$. □

Example 4.9. Solve, if possible, the matrix equation $\mathbf{Ax} = \mathbf{b}$ if

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 6 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

Solution. First note that the unknown vector $\mathbf{x}$ is in $\mathbb{R}^3$ because $\mathbf{A}$ has $n = 3$ columns. The linear system $\mathbf{Ax} = \mathbf{b}$ has $m = 2$ equations and $n = 3$ unknowns. The coefficient matrix $\mathbf{A}$ has rank $r = 2$, and therefore the solution set will contain $d = n - r = 1$ parameter. The augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ is

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 & 2 \\ 0 & 3 & 6 & -1 \end{bmatrix}.$$

Let $x_3 = t$ be the parameter and use the last row to solve for x_2:

$$x_2 = -\tfrac{1}{3} - 2t$$

Now use the first row to solve for x_1:

$$x_1 = 2 + x_2 - 2x_3 = 2 + (-\tfrac{1}{3} - 2t) - 2t = \tfrac{5}{3} - 4t.$$

Thus, the solution set to the linear system is

$$\begin{aligned} x_1 &= \tfrac{5}{3} - 4t \\ x_2 &= -\tfrac{1}{3} - 2t \\ x_3 &= t \end{aligned}$$

where t is an arbitrary number. Therefore, the matrix equation $\mathbf{Ax} = \mathbf{b}$ has an infinite number of solutions and they can all be written as

$$\mathbf{x} = \begin{bmatrix} \frac{5}{3} - 4t \\ -\frac{1}{3} - 2t \\ t \end{bmatrix}$$

www.RanaMaths.com

where t is an arbitrary number. Equivalently, $\mathbf{b}$ can be written as a linear combination of the columns of $\mathbf{A}$ in infinitely many ways. For example, choosing $t = -1$ gives the particular solution

$$\mathbf{x} = \begin{bmatrix} 17/3 \\ -7/3 \\ -1 \end{bmatrix}$$

and you can verify that

$$\mathbf{A} \begin{bmatrix} 17/3 \\ -7/3 \\ -1 \end{bmatrix} = \mathbf{b}.$$

□

Recall from Definition 3.7 that the span of a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, which we denoted by $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, is the space of vectors that can be written as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$.

Example 4.10. Is the vector $\mathbf{b}$ in the span of the vectors $\mathbf{v}_1, \mathbf{v}_2$?

$$\mathbf{b} = \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}, \ \mathbf{v}_1 = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} -5 \\ 6 \\ 1 \end{bmatrix}$$

Solution. The vector $\mathbf{b}$ is in $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ if we can find scalars x_1, x_2 such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \mathbf{b}.$$

If we let $\mathbf{A} \in \mathbb{R}^{3\times 2}$ be the matrix

$$\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2] = \begin{bmatrix} 3 & -5 \\ -2 & 6 \\ 1 & 1 \end{bmatrix}$$

then we need to solve the matrix equation $\mathbf{Ax} = \mathbf{b}$. Note that here $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$. Performing row reduction on the augmented matrix $[\mathbf{A} \ \mathbf{b}]$ we get that

$$\begin{bmatrix} 3 & -5 & 0 \\ -2 & 6 & 4 \\ 1 & 1 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2.5 \\ 0 & 1 & 1.5 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore, the linear system is consistent and has solution

$$\mathbf{x} = \begin{bmatrix} 2.5 \\ 1.5 \end{bmatrix}$$

Therefore, $\mathbf{b}$ is in $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, and $\mathbf{b}$ can be written in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$ as

$$2.5\mathbf{v}_1 + 1.5\mathbf{v}_2 = \mathbf{b}$$

□

www.RanaMaths.com

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are vectors in $\mathbb{R}^n$ and it happens to be true that $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} = \mathbb{R}^n$ then we would say that the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ **spans** all of $\mathbb{R}^n$. From Theorem 4.6, we have the following.

Theorem 4.11: Let $\mathbf{A} \in M_{m\times n}$ be a matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, that is, $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$. The following are equivalent:

(a) $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \mathbb{R}^m$
(b) Every $\mathbf{b} \in \mathbb{R}^m$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.
(c) The matrix equation $\mathbf{Ax} = \mathbf{b}$ has a solution for *any* $\mathbf{b} \in \mathbb{R}^m$.
(d) The rank of $\mathbf{A}$ is m.

Example 4.12. Do the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ span $\mathbb{R}^3$?

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$$

Solution. From Theorem 4.11, the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ span $\mathbb{R}^3$ if the matrix $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}$ has rank $r = 3$ (leading entries in its REF/RREF). The RREF of $\mathbf{A}$ is

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

which does indeed have $r = 3$ leading entries. Therefore, regardless of the choice of $\mathbf{b} \in \mathbb{R}^3$, the augmented matrix $[\mathbf{A}\ \mathbf{b}]$ will be consistent. Therefore, the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ span $\mathbb{R}^3$:

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \mathbb{R}^3.$$

In other words, every vector $\mathbf{b} \in \mathbb{R}^3$ can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. □

After this lecture you should know the following:

- how to multiply a matrix $\mathbf{A}$ with a vector $\mathbf{x}$
- that the product $\mathbf{Ax}$ is a linear combination of the columns of $\mathbf{A}$
- how to solve the matrix equation $\mathbf{Ax} = \mathbf{b}$ if $\mathbf{A}$ and $\mathbf{b}$ are known
- how to determine if a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^m$ spans all of $\mathbb{R}^m$
- the relationship between the equation $\mathbf{Ax} = \mathbf{b}$, when $\mathbf{b}$ can be written as a linear combination of the columns of $\mathbf{A}$, and when the augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ is consistent (Theorem 4.6)
- when the columns of a matrix $\mathbf{A} \in M_{m\times n}$ span all of $\mathbb{R}^m$ (Theorem 4.11)
- the basic properties of matrix-vector multiplication Theorem 4.3

www.RanaMaths.com

www.RanaMaths.com

Lecture 5

Homogeneous and Nonhomogeneous Systems

5.1 Homogeneous linear systems

We begin with a definition.

Definition 5.1: A linear system of the form $\mathbf{Ax} = \mathbf{0}$ is called a **homogeneous** linear system.

A homogeneous system $\mathbf{Ax} = \mathbf{0}$ always has at least one solution, namely, the zero solution because $\mathbf{A0} = \mathbf{0}$. A homogeneous system is therefore always consistent. The zero solution $\mathbf{x} = \mathbf{0}$ is called the **trivial solution** and any non-zero solution is called a **nontrivial solution**. From the existence and uniqueness theorem (Theorem 2.5), we know that a consistent linear system will have either one solution or infinitely many solutions. Therefore, a homogeneous linear system has nontrivial solutions if and only if its solution set has at least one parameter.

Recall that the number of parameters in the solution set is $d = n - r$, where r is the rank of the coefficient matrix $\mathbf{A}$ and n is the number of unknowns.

Example 5.2. Does the linear homogeneous system have any nontrivial solutions?

$$\begin{aligned} 3x_1 + x_2 - 9x_3 &= 0 \\ x_1 + x_2 - 5x_3 &= 0 \\ 2x_1 + x_2 - 7x_3 &= 0 \end{aligned}$$

Solution. The linear system will have a nontrivial solution if the solution set has at least one free parameter. Form the augmented matrix:

$$\begin{bmatrix} 3 & 1 & -9 & 0 \\ 1 & 1 & -5 & 0 \\ 2 & 1 & -7 & 0 \end{bmatrix}$$

www.RanaMaths.com

The RREF is:

$$\begin{bmatrix} 3 & 1 & -9 & 0 \\ 1 & 1 & -5 & 0 \\ 2 & 1 & -7 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The system is consistent. The rank of the coefficient matrix is $r = 2$ and thus there will be $d = 3 - 2 = 1$ free parameter in the solution set. If we let x_3 be the free parameter, say $x_3 = t$, then from the row equivalent augmented matrix

$$\begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we obtain that $x_2 = 3x_3 = 3t$ and $x_1 = 2x_3 = 2t$. Therefore, the general solution of the linear system is

$$x_1 = 2t$$
$$x_2 = 3t$$
$$x_3 = t$$

The general solution can be written in vector notation as

$$\mathbf{x} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} t$$

Or more compactly if we let $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$ then $\mathbf{x} = \mathbf{v}t$. Hence, any solution $\mathbf{x}$ to the linear system can be written as a linear combination of the vector $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$. In other words, the solution set of the linear system is the span of the vector $\mathbf{v}$:

$$\text{span}\{\mathbf{v}\}.$$

□

Notice that in the previous example, when solving a homogeneous system $\mathbf{Ax} = \mathbf{0}$ using row reduction, the last column of the augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{0} \end{bmatrix}$ remains unchanged (always $\mathbf{0}$) after every elementary row operation. Hence, to solve a homogeneous system, we can row reduce the coefficient matrix $\mathbf{A}$ only and then set all rows equal to zero when performing back substitution.

Example 5.3. Find the general solution of the homogenous system $\mathbf{Ax} = \mathbf{0}$ where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 1 & 4 \\ 3 & 7 & 7 & 3 & 13 \\ 2 & 5 & 5 & 2 & 9 \end{bmatrix}.$$

www.RanaMaths.com

Solution. After row reducing we obtain

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 2 & 1 & 4 \\ 3 & 7 & 7 & 3 & 13 \\ 2 & 5 & 5 & 2 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Here $n = 5$, and $r = 2$, and therefore the number of parameters in the solution set is $d = n - r = 3$. The second row of $\texttt{rref}(\mathbf{A})$ gives the equation

$$x_2 + x_3 + x_5 = 0.$$

Setting $x_5 = t_1$ and $x_3 = t_2$ as free parameters we obtain that

$$x_2 = -x_3 - x_5 = -t_2 - t_1.$$

From the first row we obtain the equation

$$x_1 + x_4 + 2x_5 = 0$$

The unknown x_5 has already been assigned, so we must now choose either x_1 or x_4 to be a parameter. Choosing $x_4 = t_3$ we obtain that

$$x_1 = -x_4 - 2x_5 = -t_3 - 2t_1$$

In summary, the general solution can be written as

$$\mathbf{x} = \begin{bmatrix} -t_3 - 2t_1 \\ -t_2 - t_1 \\ t_2 \\ t_3 \\ t_1 \end{bmatrix} = t_1 \underbrace{\begin{bmatrix} -2 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_1} + t_2 \underbrace{\begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{v}_2} + t_3 \underbrace{\begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_3} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + t_3\mathbf{v}_3$$

where t_1, t_2, t_3 are arbitrary parameters. In other words, any solution $\mathbf{x}$ is in the span of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$:

$$\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}.$$

□

The form of the general solution in Example 5.3 holds in general and is summarized in the following theorem.

Theorem 5.4: Consider the homogenous linear system $\mathbf{Ax} = \mathbf{0}$, where $\mathbf{A} \in M_{m\times n}$ and $\mathbf{0} \in \mathbb{R}^m$. Let r be the rank of $\mathbf{A}$.

1. If $r = n$ then the only solution to the system is the trivial solution $\mathbf{x} = \mathbf{0}$.

2. Otherwise, if $r < n$ and we set $d = n - r$, then there exist vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$ such that any solution $\mathbf{x}$ of the linear system can be written as

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_p\mathbf{v}_d.$$

www.RanaMaths.com

In other words, any solution $\mathbf{x}$ is in the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$:

$$\mathbf{x} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d\}.$$

A solution $\mathbf{x}$ to a homogeneous system written in the form

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_p\mathbf{v}_d$$

is said to be in **parametric vector form**.

5.2 Nonhomogeneous systems

As we have seen, a homogeneous system $\mathbf{Ax} = \mathbf{0}$ is always consistent. However, if $\mathbf{b}$ is non-zero, then the **nonhomogeneous** linear system $\mathbf{Ax} = \mathbf{b}$ may or may not have a solution. A natural question arises: What is the relationship between the solution set of the homogeneous system $\mathbf{Ax} = \mathbf{0}$ and that of the nonhomogeneous system $\mathbf{Ax} = \mathbf{b}$ when it is consistent? To answer this question, suppose that $\mathbf{p}$ is a solution to the nonhomogeneous system $\mathbf{Ax} = \mathbf{b}$, that is, $\mathbf{Ap} = \mathbf{b}$. And suppose that $\mathbf{v}$ is a solution to the homogeneous system $\mathbf{Ax} = \mathbf{0}$, that is, $\mathbf{Av} = \mathbf{0}$. Now let $\mathbf{q} = \mathbf{p} + \mathbf{v}$. Then

$$\begin{aligned}\mathbf{Aq} &= \mathbf{A}(\mathbf{p} + \mathbf{v})\\ &= \mathbf{Ap} + \mathbf{Av}\\ &= \mathbf{b} + \mathbf{0}\\ &= \mathbf{b}.\end{aligned}$$

Therefore, $\mathbf{Aq} = \mathbf{b}$. In other words, $\mathbf{q} = \mathbf{p} + \mathbf{v}$ is also a solution of $\mathbf{Ax} = \mathbf{b}$. We have therefore proved the following theorem.

Theorem 5.5: Suppose that the linear system $\mathbf{Ax} = \mathbf{b}$ is consistent and let $\mathbf{p}$ be a solution. Then any other solution $\mathbf{q}$ of the system $\mathbf{Ax} = \mathbf{b}$ can be written in the form $\mathbf{q} = \mathbf{p} + \mathbf{v}$, for some vector $\mathbf{v}$ that is a solution to the homogeneous system $\mathbf{Ax} = \mathbf{0}$.

Another way of stating Theorem 5.5 is the following: If the linear system $\mathbf{Ax} = \mathbf{b}$ is consistent and has solutions $\mathbf{p}$ and $\mathbf{q}$, then the vector $\mathbf{v} = \mathbf{q} - \mathbf{p}$ is a solution to the homogeneous system $\mathbf{Ax} = \mathbf{0}$. The proof is a simple computation:

$$\mathbf{Av} = \mathbf{A}(\mathbf{q} - \mathbf{p}) = \mathbf{Aq} - \mathbf{Ap} = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

More generally, any solution of $\mathbf{Ax} = \mathbf{b}$ can be written in the form

$$\mathbf{q} = \mathbf{p} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_p\mathbf{v}_d$$

where $\mathbf{p}$ is one particular solution of $\mathbf{Ax} = \mathbf{b}$ and the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$ span the solution set of the homogeneous system $\mathbf{Ax} = \mathbf{0}$.

www.RanaMaths.com

There is a useful geometric interpretation of the solution set of a general linear system. We saw in Lecture 3 that we can interpret the span of a set of vectors as a plane containing the zero vector $\mathbf{0}$. Now, the general solution of $\mathbf{Ax} = \mathbf{b}$ can be written as

$$\mathbf{x} = \mathbf{p} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_p\mathbf{v}_d.$$

Therefore, the solution set of $\mathbf{Ax} = \mathbf{b}$ is a shift of the span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d\}$ by the vector $\mathbf{p}$. This is illustrated in Figure 5.1.

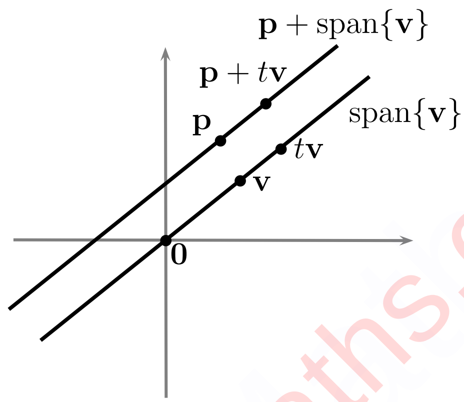


Figure 5.1: The solution sets of a homogeneous and nonhomogeneous system.

Example 5.6. Write the general solution, in parametric vector form, of the linear system

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

Solution. The RREF of the augmented matrix is:

$$\begin{bmatrix} 3 & 1 & -9 & 2 \\ 1 & 1 & -5 & 0 \\ 2 & 1 & -7 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The system is consistent and the rank of the coefficient matrix is $r = 2$. Therefore, there are $d = 3 - 2 = 1$ parameters in the solution set. Letting $x_3 = t$ be the parameter, from the second row of the RREF we have

$$x_2 = 3t - 1$$

And from the first row of the RREF we have

$$x_1 = 2t + 1$$

Therefore, the general solution of the system in parametric vector form is

$$\mathbf{x} = \begin{bmatrix} 2t+1 \\ 3t-1 \\ t \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}_{\mathbf{p}} + t \underbrace{\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}}_{\mathbf{v}}$$

www.RanaMaths.com

You should check that $\mathbf{p} = (1, -1, 0)$ solves the linear system $\mathbf{Ax} = \mathbf{b}$, and that $\mathbf{v} = (2, 3, 1)$ solves the homogeneous system $\mathbf{Ax} = \mathbf{0}$. $\square$

Example 5.7. Write the general solution, in parametric vector form, of the linear system represented by the augmented matrix

$$\begin{bmatrix} 3 & -3 & 6 & 3 \\ -1 & 1 & -2 & -1 \\ 2 & -2 & 4 & 2 \end{bmatrix}.$$

Solution. Write the general solution, in parametric vector form, of the linear system represented by the augmented matrix

$$\begin{bmatrix} 3 & -3 & 6 & 3 \\ -1 & 1 & -2 & -1 \\ 2 & -2 & 4 & 2 \end{bmatrix}$$

The RREF of the augmented matrix is

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

Here $n = 3$, $r = 1$ and therefore the solution set will have $d = 2$ parameters. Let $x_3 = t_1$ and $x_2 = t_2$. Then from the first row we obtain

$$x_1 = 1 + x_2 - 2x_3 = 1 + t_2 - 2t_1$$

The general solution in parametric vector form is therefore

$$\mathbf{x} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{p}} + t_1 \underbrace{\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_1} + t_2 \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_2}$$

You should verify that $\mathbf{p}$ is a solution to the linear system $\mathbf{Ax} = \mathbf{b}$:

$$\mathbf{Ap} = \mathbf{b}$$

And that $\mathbf{v}_1$ and $\mathbf{v}_2$ are solutions to the homogeneous linear system $\mathbf{Ax} = \mathbf{0}$:

$$\mathbf{Av}_1 = \mathbf{Av}_2 = \mathbf{0}$$

$\square$

www.RanaMaths.com

5.3 Summary

The material in this lecture is so important that we will summarize the main results. The solution set of a linear system $\mathbf{Ax} = \mathbf{b}$ can be written in the form

$$\mathbf{x} = \mathbf{p} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d$$

where $\mathbf{Ap} = \mathbf{b}$ and where each of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$ satisfies $\mathbf{Av}_i = \mathbf{0}$. Loosely speaking,

$$\{\text{Solution set of } \mathbf{Ax} = \mathbf{b}\} = \mathbf{p} + \{\text{Solution set of } \mathbf{Ax} = \mathbf{0}\}$$

or

$$\{\text{Solution set of } \mathbf{Ax} = \mathbf{b}\} = \mathbf{p} + \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d\}$$

where $\mathbf{p}$ satisfies $\mathbf{Ap} = \mathbf{b}$ and $\mathbf{Av}_i = 0$.

After this lecture you should know the following:

- what a homogeneous/nonhomogenous linear system is
- when a homogeneous linear system has nontrivial solutions
- how to write the general solution set of a homogeneous system in parametric vector form Theorem 5.4)
- how to write the solution set of a nonhomogeneous system in parametric vector form Theorem 5.5)
- the relationship between the solution sets of the nonhomogeneous equation $\mathbf{Ax} = \mathbf{b}$ and the homogeneous equation $\mathbf{Ax} = \mathbf{0}$

www.RanaMaths.com

Lecture 6

Linear Independence

6.1 Linear independence

In Lecture 3, we defined the span of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ as the collection of all possible linear combinations

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n$$

and we denoted this set as $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Thus, if $\mathbf{x} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ then by definition there exists scalars $t_1, t_2, \ldots, t_n$ such that

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n.$$

A natural question that arises is whether or not there are multiple ways to express $\mathbf{x}$ as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. For example, if $\mathbf{v}_1 = (1, 2)$, $\mathbf{v}_2 = (0, 1)$, $\mathbf{v}_3 = (-1, -1)$, and $\mathbf{x} = (3, -1)$ then you can verify that $\mathbf{x} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\mathbf{x}$ can be written in infinitely many ways using $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$. Here are three ways:

$$\begin{aligned}
\mathbf{x} &= 3\mathbf{v}_1 - 7\mathbf{v}_2 + 0\mathbf{v}_3 \\
\mathbf{x} &= -4\mathbf{v}_1 + 0\mathbf{v}_2 - 7\mathbf{v}_3 \\
\mathbf{x} &= 0\mathbf{v}_1 - 4\mathbf{v}_2 - 3\mathbf{v}_3.
\end{aligned}$$

The fact that $\mathbf{x}$ can be written in more than one way in terms of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ suggests that there might be a redundancy in the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. In fact, it is not hard to see that $\mathbf{v}_3 = -\mathbf{v}_1 + \mathbf{v}_2$, and thus $\mathbf{v}_3 \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. The preceding discussion motivates the following definition.

Definition 6.1: A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is said to be **linearly dependent** if some $\mathbf{v}_j$ can be written as a linear combination of the other vectors, that is, if

$$\mathbf{v}_j \in \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n\}.$$

If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is not linearly dependent then we say that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is **linearly independent**.

www.RanaMaths.com

Example 6.2. Consider the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Show that they are **linearly dependent**.

Solution. By inspection, we have

$$2\mathbf{v}_1 + \mathbf{v}_3 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \mathbf{v}_2$$

Thus, $\mathbf{v}_2 \in \text{span}\{\mathbf{v}_1, \mathbf{v}_3\}$ and therefore $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is **linearly dependent**. □

Notice that in the previous example, the equation $2\mathbf{v}_1 + \mathbf{v}_3 = \mathbf{v}_2$ is equivalent to

$$2\mathbf{v}_1 - \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}.$$

Hence, because $\{\mathbf{v}_1, \mathbf{v}_2\, \mathbf{v}_3\}$ is a linearly dependent set, it is possible to write the zero vector $\mathbf{0}$ as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2\, \mathbf{v}_3\}$ where **not all the coefficients in the linear combination are zero**. This leads to the following characterization of linear independence.

Theorem 6.3: The set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent if and only if $\mathbf{0}$ can be written in only one way as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. In other words, if

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n = \mathbf{0}$$

then necessarily the coefficients $t_1, t_2, \ldots, t_n$ are all zero.

Proof. If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent then every vector $\mathbf{x} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ can be written uniquely as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, and this applies to the particular case of the zero vector $\mathbf{x} = \mathbf{0}$.

Now assume that $\mathbf{0}$ can be written uniquely as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. In other words, assume that if

$$t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n = \mathbf{0}$$

then $t_1 = t_2 = \cdots = t_n = 0$. Now take any $\mathbf{x} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and suppose that there are two ways to write $\mathbf{x}$ in terms of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$:

$$r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n = \mathbf{x}$$
$$s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n = \mathbf{x}.$$

Subtracting the second equation from the first we obtain that

$$(r_1 - s_1)\mathbf{v}_1 + (r_2 - s_2)\mathbf{v}_2 + \cdots + (r_n - s_n)\mathbf{v}_n = \mathbf{x} - \mathbf{x} = \mathbf{0}.$$

www.RanaMaths.com

The above equation is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ resulting in the zero vector $\mathbf{0}$. But we are assuming that the only way to write $\mathbf{0}$ in terms of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is if all the coefficients are zero. Therefore, we must have $r_1 - s_1 = 0$, $r_2 - s_2 = 0, \ldots, r_n - s_n = 0$, or equivalently that $r_1 = s_1, r_2 = s_2, \ldots, r_n = s_n$. Therefore, the linear combinations

$$r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n = \mathbf{x}$$
$$s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_n\mathbf{v}_n = \mathbf{x}$$

are actually the same. Therefore, each $\mathbf{x} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ can be written uniquely in terms of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, and thus $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a linearly independent set. □

Because of Theorem 6.3, an alternative definition of linear independence of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is that the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{0}$$

has only the trivial solution, i.e., the solution $x_1 = x_2 = \cdots = x_n = 0$. Thus, if $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly dependent, then there exist scalars $x_1, x_2, \ldots, x_n$ not all zero such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{0}.$$

Hence, if we suppose for instance that $x_n \neq 0$ then we can write $\mathbf{v}_n$ in terms of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ as follows:

$$\mathbf{v}_n = -\tfrac{x_1}{x_n}\mathbf{v}_1 - \tfrac{x_2}{x_n}\mathbf{v}_2 - \cdots - \tfrac{x_{n-1}}{x_n}\mathbf{v}_{n-1}.$$

In other words, $\mathbf{v}_n \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}\}$.

According to Theorem 6.3, the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent if the equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{0} \tag{6.1}$$

has only the trivial solution. Now, the vector equation (6.1) is a homogeneous linear system of equations with coefficient matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

Therefore, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent if and only if the the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution. But the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution if there are no free parameters in its solution set. We therefore have the following.

Theorem 6.4: The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent if and only if the the rank of $\mathbf{A}$ is $r = n$, that is, if the number of leading entries r in the REF (or RREF) of $\mathbf{A}$ is exactly n.

Example 6.5. Are the vectors below linearly independent?

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 5 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 8 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}$$

www.RanaMaths.com

Solution. Let $\mathbf{A}$ be the matrix

$$\mathbf{A} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} = \begin{bmatrix}0 & 1 & 4\\ 1 & 2 & -1\\ 5 & 8 & 0\end{bmatrix}$$

Performing elementary row operations we obtain

$$\mathbf{A} \sim \begin{bmatrix}1 & 2 & -1\\ 0 & 1 & 4\\ 0 & 0 & 13\end{bmatrix}$$

Clearly, $r = \text{rank}(\mathbf{A}) = 3$, which is equal to the number of vectors $n = 3$. Therefore, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent. □

Example 6.6. Are the vectors below linearly independent?

$$\mathbf{v}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix}4\\5\\6\end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix}2\\1\\0\end{bmatrix}$$

Solution. Let $\mathbf{A}$ be the matrix

$$\mathbf{A} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} = \begin{bmatrix}1 & 4 & 2\\ 2 & 5 & 1\\ 3 & 6 & 0\end{bmatrix}$$

Performing elementary row operations we obtain

$$\mathbf{A} \sim \begin{bmatrix}1 & 4 & 2\\ 0 & -3 & -3\\ 0 & 0 & 0\end{bmatrix}$$

Clearly, $r = \text{rank}(\mathbf{A}) = 2$, which is **not** equal to the number of vectors, $n = 3$. Therefore, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent. We will find a nontrivial linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ that gives the zero vector $\mathbf{0}$. The REF of $\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$ is

$$\mathbf{A} \sim \begin{bmatrix}1 & 4 & 2\\ 0 & -3 & -3\\ 0 & 0 & 0\end{bmatrix}$$

Since $r = 2$, the solution set of the linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has $d = n - r = 1$ free parameter. Using back substitution on the REF above, we find that the general solution of $\mathbf{A}\mathbf{x} = \mathbf{0}$ written in parametric form is

$$\mathbf{x} = t\begin{bmatrix}2\\-1\\1\end{bmatrix}$$

The vector

$$\mathbf{v} = \begin{bmatrix}2\\-1\\1\end{bmatrix}$$

www.RanaMaths.com

spans the solution set of the system $\mathbf{Ax} = \mathbf{0}$. Choosing for instance $t = 2$ we obtain the solution

$$\mathbf{x} = t \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix}.$$

Therefore,

$$4\mathbf{v}_1 - 2\mathbf{v}_2 + 2\mathbf{v}_3 = \mathbf{0}$$

is a non-trivial linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ that gives the zero vector $\mathbf{0}$. And, for instance,

$$\mathbf{v}_3 = -2\mathbf{v}_1 + \mathbf{v}_2$$

that is, $\mathbf{v}_3 \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. □

Below we record some simple observations on the linear independence of simple sets:

- A set consisting of a single non-zero vector $\{\mathbf{v}_1\}$ is linearly independent. Indeed, if $\mathbf{v}_1$ is non-zero then

$$t\mathbf{v}_1 = \mathbf{0}$$

  is true if and only if $t = 0$.

- A set consisting of two non-zero vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent if and only if neither of the vectors is a multiple of the other. For example, if $\mathbf{v}_2 = t\mathbf{v}_1$ then

$$t\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$$

  is a non-trivial linear combination of $\mathbf{v}_1, \mathbf{v}_2$ giving the zero vector $\mathbf{0}$.

- Any set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ containing the zero vector, say that $\mathbf{v}_p = \mathbf{0}$, is linearly dependent. For example, the linear combination

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_{p-1} + 2\mathbf{v}_p = \mathbf{0}$$

  is a non-trivial linear combination giving the zero vector $\mathbf{0}$.

6.2 The maximum size of a linearly independent set

The next theorem puts a constraint on the maximum size of a linearly independent set in $\mathbb{R}^n$.

Theorem 6.7: Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a set of vectors in $\mathbb{R}^n$. If $p > n$ then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are linearly dependent. Equivalently, if the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in $\mathbb{R}^n$ are linearly independent then $p \leq n$.

www.RanaMaths.com

Proof. Let $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p \end{bmatrix}$. Thus, $\mathbf{A}$ is a $n \times p$ matrix. Since $\mathbf{A}$ has n rows, the maximum rank of $\mathbf{A}$ is n, that is $r \leq n$. Therefore, the number of free parameters $d = p - r$ is always positive because $p > n \geq r$. Thus, the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has non-trivial solutions. In other words, there is some non-zero vector $\mathbf{x} \in \mathbb{R}^p$ such that

$$\mathbf{A}\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

and therefore $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is linearly dependent. □

Theorem 6.7 will be used when we discuss the notion of the **dimension** of a space. Although we have not discussed the meaning of dimension, the above theorem says that in n-dimensional space $\mathbb{R}^n$, a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ consisting of more than n vectors is automatically linearly dependent.

Example 6.8. Are the vectors below linearly independent?

$$\mathbf{v}_1 = \begin{bmatrix} 8 \\ 3 \\ 0 \\ -2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ 11 \\ -4 \\ 6 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} 3 \\ -9 \\ -5 \\ 3 \end{bmatrix}, \mathbf{v}_5 = \begin{bmatrix} 0 \\ -2 \\ -7 \\ 7 \end{bmatrix}.$$

Solution. The vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5$ are in $\mathbb{R}^4$. Therefore, by Theorem 6.7, the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_5\}$ is linearly dependent. To see this explicitly, let $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 & \mathbf{v}_5 \end{bmatrix}$. Then

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

One solution to the linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = (-1, 1, 0, -2, -1)$ and therefore

$$(-1)\mathbf{v}_1 + (1)\mathbf{v}_2 + (0)\mathbf{v}_3 + (-2)\mathbf{v}_4 + (-1)\mathbf{v}_5 = \mathbf{0}$$

□

Example 6.9. Suppose that the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent. Show that the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is also linearly independent.

Solution. We must argue that if there exists scalars x_1, x_2, x_3 such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$$

then necessarily x_1, x_2, x_3 are all zero. Suppose then that there exists scalars x_1, x_2, x_3 such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}.$$

Then clearly it holds that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 + 0\mathbf{v}_4 = \mathbf{0}.$$

But the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent, and therefore, it is necessary that x_1, x_2, x_3 are all zero. This proves that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are also linearly independent. □

www.RanaMaths.com

The previous example can be generalized as follows: If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d\}$ is linearly independent then any (non-empty) subset of the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d\}$ is also linearly independent.

After this lecture you should know the following:

- the definition of linear independence and be able to explain it to a colleague
- how to test if a given set of vectors are linearly independent (Theorem 6.4)
- the relationship between the linear independence of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ and the solution set of the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$, where $\mathbf{A} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p\end{bmatrix}$
- that in $\mathbb{R}^n$, any set of vectors consisting of more than n vectors is automatically linearly dependent (Theorem 6.7)

www.RanaMaths.com

Lecture 7

Introduction to Linear Mappings

7.1 Vector mappings

By a **vector mapping** we mean simply a function

$$\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m.$$

The **domain** of T is $\mathbb{R}^n$ and the **co-domain** of T is $\mathbb{R}^m$. The case $n = m$ is allowed of course. In engineering or physics, the domain is sometimes called the **input space** and the co-domain is called the **output space**. Using this terminology, the points $\mathbf{x}$ in the domain are called the **inputs** and the points $\mathsf{T}(\mathbf{x})$ produced by the mapping are called the **outputs**.

Definition 7.1: The vector $\mathbf{b} \in \mathbb{R}^m$ is in the **range** of T, or in the **image** of T, if there exists some $\mathbf{x} \in \mathbb{R}^n$ such that $\mathsf{T}(\mathbf{x}) = \mathbf{b}$.

In other words, $\mathbf{b}$ is in the range of T if there is an input $\mathbf{x}$ in the domain of T that outputs $\mathbf{b} = \mathsf{T}(\mathbf{x})$. In general, not every point in the co-domain of T is in the range of T. For example, consider the vector mapping $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$\mathsf{T}(\mathbf{x}) = \begin{bmatrix} x_1^2 \sin(x_2) - \cos(x_1^2 - 1) \\ x_1^2 + x_2^2 + 1 \end{bmatrix}.$$

The vector $\mathbf{b} = (3, -1)$ is not in the range of T because the second component of $\mathsf{T}(\mathbf{x})$ is positive. On the other hand, $\mathbf{b} = (-1, 2)$ is in the range of T because

$$\mathsf{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1^2 \sin(0) - \cos(1^2 - 1) \\ 1^2 + 0^2 + 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \mathbf{b}.$$

Hence, a corresponding input for this particular $\mathbf{b}$ is $\mathbf{x} = (1, 0)$. In Figure 7.1 we illustrate the general setup of how the domain, co-domain, and range of a mapping are related. A crucial idea is that the range of T may not equal the co-domain.

www.RanaMaths.com

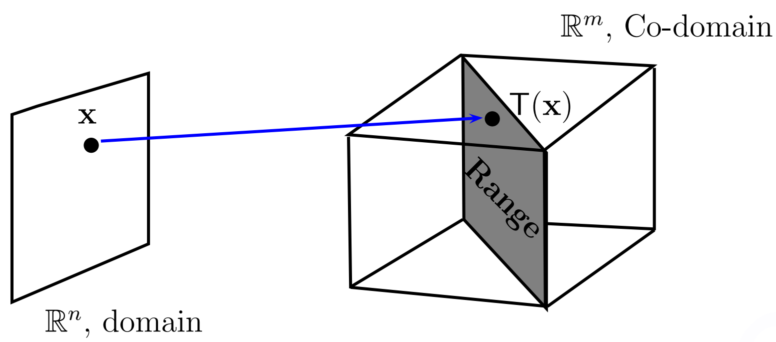


Figure 7.1: The domain, co-domain, and range of a mapping.

7.2 Linear mappings

For our purposes, vector mappings $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ can be organized into two categories: (1) linear mappings and (2) nonlinear mappings.

Definition 7.2: The vector mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **linear** if the following conditions hold:

- For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, it holds that $\mathsf{T}(\mathbf{u} + \mathbf{v}) = \mathsf{T}(\mathbf{u}) + \mathsf{T}(\mathbf{v})$.
- For any $\mathbf{u} \in \mathbb{R}^n$ and any scalar c, it holds that $\mathsf{T}(c\mathbf{u}) = c\mathsf{T}(\mathbf{u})$.

If T is not linear then it is said to be **nonlinear**.

As an example, the mapping

$$\mathsf{T}(\mathbf{x}) = \begin{bmatrix} x_1^2 \sin(x_2) - \cos(x_1^2 - 1) \\ x_1^2 + x_2^2 + 1 \end{bmatrix}$$

is **nonlinear**. To see this, previously we computed that

$$\mathsf{T}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

www.RanaMaths.com

If T were linear then by property (2) of Definition 7.2 the following must hold:

$$\begin{aligned} \mathsf{T}\left(\begin{bmatrix}3\\0\end{bmatrix}\right) &= \mathsf{T}\left(3\begin{bmatrix}1\\0\end{bmatrix}\right) \\ &= 3\mathsf{T}\left(\begin{bmatrix}1\\0\end{bmatrix}\right) \\ &= 3\begin{bmatrix}-1\\2\end{bmatrix} \\ &= \begin{bmatrix}-3\\6\end{bmatrix}. \end{aligned}$$

However,

$$\begin{aligned} \mathsf{T}\left(\begin{bmatrix}3\\0\end{bmatrix}\right) &= \begin{bmatrix}3^2\sin(0)-\cos(3^2-1)\\ 3^2+0^2+1\end{bmatrix} \\ &= \begin{bmatrix}-\cos(8)\\10\end{bmatrix} \\ &\neq \begin{bmatrix}-3\\6\end{bmatrix}. \end{aligned}$$

Example 7.3. Is the vector mapping $\mathsf{T}:\mathbb{R}^2\to\mathbb{R}^3$ linear?

$$\mathsf{T}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}2x_1-x_2\\x_1+x_2\\-x_1-3x_2\end{bmatrix}$$

Solution. We must verify that the two conditions in Definition 7.2 hold. For the first condition, take arbitrary vectors $\mathbf{u}=(u_1,u_2)$ and $\mathbf{v}=(v_1,v_2)$. We compute:

$$\begin{aligned} \mathsf{T}(\mathbf{u}+\mathbf{v}) &= \mathsf{T}\left(\begin{bmatrix}u_1+v_1\\u_2+v_2\end{bmatrix}\right) \\ &= \begin{bmatrix}2(u_1+v_1)-(u_2+v_2)\\(u_1+v_1)+(u_2+v_2)\\-(u_1+v_1)-3(u_2+v_2)\end{bmatrix} \\ &= \begin{bmatrix}2u_1+2v_1-u_2-v_2\\u_1+v_1+u_2+v_2\\-u_1-v_1-3u_2-3v_2\end{bmatrix} \\ &= \begin{bmatrix}2u_1-u_2+2v_1-v_2\\u_1+u_2+v_1+v_2\\-u_1-3u_2-v_1-3v_2\end{bmatrix} \\ &= \begin{bmatrix}2u_1-u_2\\u_1+u_2\\-u_1-3u_2\end{bmatrix} + \begin{bmatrix}2v_1-v_2\\v_1+v_2\\-v_1-3v_2\end{bmatrix} \\ &= \mathsf{T}(\mathbf{u})+\mathsf{T}(\mathbf{v}) \end{aligned}$$

www.RanaMaths.com

Therefore, for arbitrary $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, it holds that

$$\mathsf{T}(\mathbf{u}+\mathbf{v}) = \mathsf{T}(\mathbf{u}) + \mathsf{T}(\mathbf{v}).$$

To prove the second condition, let $c \in \mathbb{R}$ be an arbitrary scalar. Then:

$$\begin{aligned}
\mathsf{T}(c\mathbf{u}) = \mathsf{T}\left(\begin{bmatrix} cu_1 \\ cu_2 \end{bmatrix}\right) &= \begin{bmatrix} 2(cu_1) - (cu_2) \\ (cu_1) + (cu_2) \\ -(cu_1) - 3(cu_2) \end{bmatrix} \\
&= \begin{bmatrix} c(2u_1 - u_2) \\ c(u_1 + u_2) \\ c(-u_1 - 3u_2) \end{bmatrix} \\
&= c \begin{bmatrix} 2u_1 - u_2 \\ u_1 + u_2 \\ -u_1 - 3u_2 \end{bmatrix} \\
&= c\mathsf{T}(\mathbf{u})
\end{aligned}$$

Therefore, both conditions of Definition 7.2 hold, and thus T is a linear map. □

Example 7.4. Let $\alpha \geq 0$ and define the mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ by the formula $\mathsf{T}(\mathbf{x}) = \alpha\mathbf{x}$. If $0 \leq \alpha \leq 1$ then T is called a **contraction** and if $\alpha > 1$ then T is called a **dilation**. In either case, show that T is a linear mapping.

Solution. Let $\mathbf{u}$ and $\mathbf{v}$ be arbitrary. Then

$$\mathsf{T}(\mathbf{u}+\mathbf{v}) = \alpha(\mathbf{u}+\mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v} = \mathsf{T}(\mathbf{u}) + \mathsf{T}(\mathbf{v}).$$

This shows that condition (1) in Definition 7.2 holds. To show that the second condition holds, let c is any number. Then

$$\mathsf{T}(c\mathbf{x}) = \alpha(c\mathbf{x}) = \alpha c\mathbf{x} = c(\alpha\mathbf{x}) = c\mathsf{T}(\mathbf{x}).$$

Therefore, both conditions of Definition 7.2 hold, and thus T is a linear mapping. To see a particular example, consider the case $\alpha = \frac{1}{2}$ and $n = 3$. Then,

$$\mathsf{T}(\mathbf{x}) = \tfrac{1}{2}\mathbf{x} = \begin{bmatrix} \frac{1}{2}x_1 \\ \frac{1}{2}x_2 \\ \frac{1}{2}x_3 \end{bmatrix}.$$

□

www.RanaMaths.com

7.3 Matrix mappings

Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and a vector $\mathbf{x} \in \mathbb{R}^n$, in Lecture 4 we defined matrix-vector multiplication between $\mathbf{A}$ and $\mathbf{x}$ as an operation that produces a new output vector $\mathbf{Ax} \in \mathbb{R}^m$. We discussed that we could interpret $\mathbf{A}$ as a mapping that takes the input vector $\mathbf{x} \in \mathbb{R}^n$ and produces the output vector $\mathbf{Ax} \in \mathbb{R}^m$. We can therefore associate to each matrix $\mathbf{A}$ a vector mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$\mathsf{T}(\mathbf{x}) = \mathbf{Ax}.$$

Such a mapping T will be called a **matrix mapping** corresponding to $\mathbf{A}$ and when convenient we will use the notation $\mathsf{T}_{\mathbf{A}}$ to indicate that $\mathsf{T}_{\mathbf{A}}$ is associated to $\mathbf{A}$. We proved in Lecture 4 (Theorem 4.3), that for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and scalar c, matrix-vector multiplication satisfies the properties:

1. $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av}$
2. $\mathbf{A}(c\mathbf{u}) = c\mathbf{Au}$.

The following theorem is therefore immediate.

Theorem 7.5: To a given matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ associate the mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ defined by the formula $\mathsf{T}(\mathbf{x}) = \mathbf{Ax}$. Then T is a linear mapping.

Example 7.6. Is the vector mapping $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^3$ linear?

$$\mathsf{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 - x_2 \\ x_1 + x_2 \\ -x_1 - 3x_2 \end{bmatrix}$$

Solution. In Example 7.3 we showed that T is a linear mapping using Definition 7.2. Alternatively, we observe that T is a mapping defined using matrix-vector multiplication because

$$\mathsf{T}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 - x_2 \\ x_1 + x_2 \\ -x_1 - 3x_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 1 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Therefore, T is a matrix mapping corresponding to the matrix

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

that is, $\mathsf{T}(\mathbf{x}) = \mathbf{Ax}$. By Theorem 7.5, T is a linear mapping. □

Let $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ be a vector mapping. Recall that $\mathbf{b} \in \mathbb{R}^m$ is in the range of T if there is some input vector $\mathbf{x} \in \mathbb{R}^n$ such that $\mathsf{T}(\mathbf{x}) = \mathbf{b}$. In this case, we say that **b is the image of x under** T or that **x is mapped to b under** T. If T is a nonlinear mapping, finding a specific vector $\mathbf{x}$ such that $\mathsf{T}(\mathbf{x}) = \mathbf{b}$ is generally a difficult problem. However, if $\mathsf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is a matrix mapping, then it is clear that finding such a vector $\mathbf{x}$ is equivalent to solving the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. In summary, we have the following theorem.

Theorem 7.7: Let $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ be a matrix mapping corresponding to $\mathbf{A}$, that is, $\mathsf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Then $\mathbf{b} \in \mathbb{R}^m$ is in the range of T if and only if the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution.

Let $\mathsf{T}_\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$ be a matrix mapping, that is, $\mathsf{T}_\mathbf{A}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. We proved that the output vector $\mathbf{A}\mathbf{x}$ is a linear combination of the columns of $\mathbf{A}$ where the coefficients in the linear combination are the components of $\mathbf{x}$. Explicitly, if $\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$ and the components of $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ then

$$\mathbf{A}\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n.$$

Therefore, the range of the matrix mapping $\mathsf{T}_\mathbf{A}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is

$$\text{Range}(\mathsf{T}_\mathbf{A}) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}.$$

In words, the range of a matrix mapping is the span of its columns. Therefore, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ span all of $\mathbb{R}^m$ then every vector $\mathbf{b} \in \mathbb{R}^m$ is in the range of $\mathsf{T}_\mathbf{A}$.

Example 7.8. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -4 \\ 1 & 5 & 2 \\ -3 & -7 & -6 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} -2 \\ 4 \\ 12 \end{bmatrix}.$$

Is the vector $\mathbf{b}$ in the range of the matrix mapping $\mathsf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x}$?

Solution. From Theorem 7.7, $\mathbf{b}$ is in the range of T if and only if the the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution. To solve the system $\mathbf{A}\mathbf{x} = \mathbf{b}$, row reduce the augmented matrix $\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}$:

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

The system is consistent and the (unique) solution is $\mathbf{x} = (-11, 3, 0)$. Therefore, $\mathbf{b}$ is in the range of T. □

7.4 Examples

If $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, then for any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ and scalars $c_1, c_2, \ldots, c_p$, it holds that

$$\mathsf{T}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_d) = c_1\mathsf{T}(\mathbf{v}_1) + c_2\mathsf{T}(\mathbf{v}_2) + \cdots + c_d\mathsf{T}(\mathbf{v}_p). \qquad (\star)$$

Therefore, if all you know are the values $\mathsf{T}(\mathbf{v}_1), \mathsf{T}(\mathbf{v}_2), \ldots, \mathsf{T}(\mathbf{v}_p)$ and T is linear, then you can compute $\mathsf{T}(\mathbf{v})$ for every

$$\mathbf{v} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}.$$

Example 7.9. Let $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{u}$ to $\mathsf{T}(\mathbf{u}) = (3, 4)$ and maps $\mathbf{v}$ to $\mathsf{T}(\mathbf{v}) = (-2, 5)$. Find $\mathsf{T}(2\mathbf{u} + 3\mathbf{v})$.

Solution. Because T is a linear mapping we have that

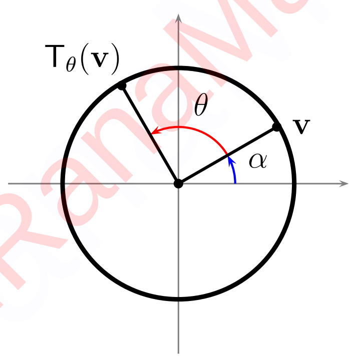
$$\mathsf{T}(2\mathbf{u} + 3\mathbf{v}) = \mathsf{T}(2\mathbf{u}) + \mathsf{T}(3\mathbf{v}) = 2\mathsf{T}(\mathbf{u}) + 3\mathsf{T}(\mathbf{v}).$$

We know that $\mathsf{T}(\mathbf{u}) = (3, 4)$ and $\mathsf{T}(\mathbf{v}) = (-2, 5)$. Therefore,

$$\mathsf{T}(2\mathbf{u} + 3\mathbf{v}) = 2\mathsf{T}(\mathbf{u}) + 3\mathsf{T}(\mathbf{v}) = 2\begin{bmatrix}3\\4\end{bmatrix} + 3\begin{bmatrix}-2\\5\end{bmatrix} = \begin{bmatrix}0\\23\end{bmatrix}.$$

□

Example 7.10. (Rotations) Let $\mathsf{T}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the mapping on the 2D plane that rotates every $\mathbf{v} \in \mathbb{R}^2$ by an angle θ. Write down a formula for T_θ and show that T_θ is a linear mapping.



Solution. If $\mathbf{v} = (\cos(\alpha), \sin(\alpha))$ then

$$\mathsf{T}_\theta(\mathbf{v}) = \begin{bmatrix}\cos(\alpha + \theta)\\ \sin(\alpha + \theta)\end{bmatrix}.$$

Then from the angle sum trigonometric identities:

$$\mathsf{T}_\theta(\mathbf{v}) = \begin{bmatrix}\cos(\alpha + \theta)\\ \sin(\alpha + \theta)\end{bmatrix} = \begin{bmatrix}\cos(\alpha)\cos(\theta) - \sin(\alpha)\sin(\theta)\\ \cos(\alpha)\sin(\theta) + \sin(\alpha)\cos(\theta)\end{bmatrix}$$

But

$$\mathsf{T}_\theta(\mathbf{v}) = \begin{bmatrix}\cos(\alpha)\cos(\theta) - \sin(\alpha)\sin(\theta)\\ \cos(\alpha)\sin(\theta) + \sin(\alpha)\cos(\theta)\end{bmatrix} = \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix}\underbrace{\begin{bmatrix}\cos(\alpha)\\ \sin(\alpha)\end{bmatrix}}_{\mathbf{v}}.$$

www.RanaMaths.com

If we scale $\mathbf{v}$ by any $c > 0$ then performing the same computation as above we obtain that $\mathsf{T}_\theta(c\mathbf{v}) = c\mathsf{T}(\mathbf{v})$. Therefore, T_θ is a matrix mapping with corresponding matrix

$$\mathbf{A} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Thus, T_θ is a linear mapping. □

Example 7.11. (Projections) Let $\mathsf{T} : \mathbb{R}^3 \to \mathbb{R}^2$ be the vector mapping

$$\mathsf{T}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}.$$

Show that T is a linear mapping and describe the range of T.

Solution. First notice that

$$\mathsf{T}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Thus, T is a matrix mapping corresponding to the matrix

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

Therefore, T is a linear mapping. Geometrically, T takes the vector $\mathbf{x}$ and projects it to the (x_1, x_2) plane, see Figure 7.2. What is the range of T? The range of T consists of all vectors in $\mathbb{R}^3$ of the form

$$\mathbf{b} = \begin{bmatrix} t \\ s \\ 0 \end{bmatrix}$$

where the numbers t and s are arbitrary. For each $\mathbf{b}$ in the range of T, there are infinitely many $\mathbf{x}$'s such that $\mathsf{T}(\mathbf{x}) = \mathbf{b}$. □

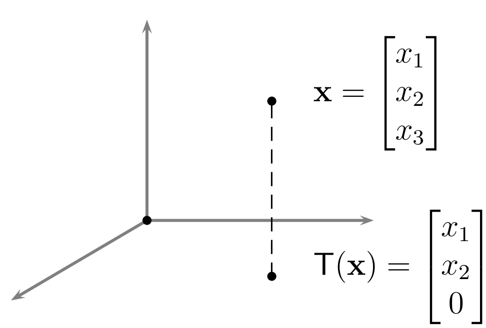


Figure 7.2: Projection onto the (x_1, x_2) plane

www.RanaMaths.com

After this lecture you should know the following:

- what a vector mapping is
- what the range of a vector mapping is
- that the co-domain and range of a vector mapping are generally not the same
- what a linear mapping is and how to check when a given mapping is linear
- what a matrix mapping is and that they are linear mappings
- how to determine if a vector $\mathbf{b}$ is in the range of a matrix mapping
- the formula for a rotation in $\mathbb{R}^2$ by an angle θ

www.RanaMaths.com

Lecture 8

Onto and One-to-One Mappings, and the Matrix of a Linear Mapping

8.1 Onto Mappings

We have seen through examples that the range of a vector mapping (linear or nonlinear) is not always the entire co-domain. For example, if $\mathsf{T}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is a matrix mapping and $\mathbf{b}$ is such that the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has no solutions then the range of T does not contain $\mathbf{b}$ and thus the range is not the whole co-domain.

Definition 8.1: A vector mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** if for each $\mathbf{b} \in \mathbb{R}^m$ there is at least one $\mathbf{x} \in \mathbb{R}^n$ such that $\mathsf{T}(\mathbf{x}) = \mathbf{b}$.

For a matrix mapping $\mathsf{T}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, the range of $\mathsf{T}_{\mathbf{A}}$ is the span of the columns of $\mathbf{A}$. Therefore:

Theorem 8.2: Let $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ be the matrix mapping $\mathsf{T}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in M_{m\times n}$. Then $\mathsf{T}_{\mathbf{A}}$ is onto if and only if the columns of $\mathbf{A}$ span all of $\mathbb{R}^m$.

Combining Theorem 4.11 and Theorem 8.2 we have:

Theorem 8.3: Let $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ be the matrix mapping $\mathsf{T}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then $\mathsf{T}_{\mathbf{A}}$ is onto if and only if $r = \operatorname{rank}(\mathbf{A}) = m$.

Example 8.4. Let $\mathsf{T} : \mathbb{R}^3 \to \mathbb{R}^3$ be the matrix mapping with corresponding matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 \\ -3 & -4 & 2 \\ 5 & 2 & 3 \end{bmatrix}$$

Is $\mathsf{T}_{\mathbf{A}}$ onto?

www.RanaMaths.com

Solution. The rref(**A**) is

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

Therefore, $r = \text{rank}(\mathbf{A}) = 3$. The dimension of the co-domain is $m = 3$ and therefore $\mathsf{T}_\mathbf{A}$ is onto. Therefore, the columns of **A** span all of $\mathbb{R}^3$, that is, every $\mathbf{b} \in \mathbb{R}^3$ can be written as a linear combination of the columns of **A**:

$$\text{span}\left\{ \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix} \right\} = \mathbb{R}^3$$

□

Example 8.5. Let $\mathsf{T}_\mathbf{A} : \mathbb{R}^4 \to \mathbb{R}^3$ be the matrix mapping with corresponding matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 & 4 \\ -1 & 4 & 1 & 8 \\ 2 & 0 & -2 & 0 \end{bmatrix}$$

Is $\mathsf{T}_\mathbf{A}$ onto?

Solution. The rref(**A**) is

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -1 & 4 \\ -1 & 4 & 1 & 8 \\ 2 & 0 & -2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore, $r = \text{rank}(\mathbf{A}) = 2$. The dimension of the co-domain is $m = 3$ and therefore $\mathsf{T}_\mathbf{A}$ is **not** onto. Notice that $\mathbf{v}_3 = -\mathbf{v}_1$ and $\mathbf{v}_4 = 2\mathbf{v}_2$. Thus, $\mathbf{v}_3$ and $\mathbf{v}_4$ are already in the span of the columns $\mathbf{v}_1, \mathbf{v}_2$. Therefore,

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \neq \mathbb{R}^3.$$

□

Below is a theorem which places restrictions on the size of the domain of an onto mapping.

Theorem 8.6: Suppose that $\mathsf{T}_\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix mapping corresponding to $\mathbf{A} \in M_{m\times n}$. If $\mathsf{T}_\mathbf{A}$ is onto then $m \leq n$.

Proof. If $\mathsf{T}_\mathbf{A}$ is onto then the rref(**A**) has $r = m$ leading 1's. Therefore, **A** has at least m columns. The number of columns of **A** is n. Therefore, $m \leq n$. □

An equivalent way of stating Theorem 8.6 is the following.

Corollary 8.7: If $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix mapping corresponding to $\mathbf{A} \in M_{m\times n}$ and $n < m$ then $\mathsf{T}_{\mathbf{A}}$ cannot be onto.

Intuitively, if the domain $\mathbb{R}^n$ is "smaller" than the co-domain $\mathbb{R}^m$ **and** $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ is linear then $\mathsf{T}_{\mathbf{A}}$ cannot be onto. For example, a matrix mapping $\mathsf{T}_{\mathbf{A}} : \mathbb{R} \to \mathbb{R}^2$ cannot be onto. Linearity plays a key role in this. In fact, there exists a continuous (nonlinear) function $f : \mathbb{R} \to \mathbb{R}^2$ whose range is a square! In this case, the domain is 1-dimensional and the range is 2-dimensional. This situation cannot happen when the mapping is linear.

Example 8.8. Let $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^3$ be the matrix mapping with corresponding matrix

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

Is $\mathsf{T}_{\mathbf{A}}$ onto?

Solution. $\mathsf{T}_{\mathbf{A}}$ is onto because the domain is $\mathbb{R}^2$ and the co-domain is $\mathbb{R}^3$. Intuitively, two vectors are not enough to span $\mathbb{R}^3$. Geometrically, two vectors in $\mathbb{R}^3$ span a 2D plane going through the origin. The vectors not on the plane span$\{\mathbf{v}_1, \mathbf{v}_2\}$ are not in the range of $\mathsf{T}_{\mathbf{A}}$. □

8.2 One-to-One Mappings

Given a linear mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$, the question of whether $\mathbf{b} \in \mathbb{R}^m$ is in the range of T is an **existence** question. Indeed, if $\mathbf{b} \in \text{Range}(\mathsf{T})$ then there **exists** a $\mathbf{x} \in \mathbb{R}^m$ such that $\mathsf{T}(\mathbf{x}) = \mathbf{b}$. We now want to look at the problem of whether $\mathbf{x}$ is **unique**. That is, does there exist a distinct $\mathbf{y}$ such that $\mathsf{T}(\mathbf{y}) = \mathbf{b}$.

Definition 8.9: A vector mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **one-to-one** if for each $\mathbf{b} \in \text{Range}(\mathsf{T})$ there exists only one $\mathbf{x} \in \mathbb{R}^n$ such that $\mathsf{T}(\mathbf{x}) = \mathbf{b}$.

When T is a linear mapping, we have all the tools necessary to give a complete description of when T is one-to-one. To do this, we use the fact that if $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is linear then $\mathsf{T}(\mathbf{0}) = \mathbf{0}$. Here is one proof: $\mathsf{T}(\mathbf{0}) = \mathsf{T}(\mathbf{x} - \mathbf{x}) = \mathsf{T}(\mathbf{x}) - \mathsf{T}(\mathbf{x}) = \mathbf{0}$.

Theorem 8.10: Let $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ be linear. Then T is one-to-one if and only if $\mathsf{T}(\mathbf{x}) = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$.

If $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix mapping then according to Theorem 8.10, $\mathsf{T}_{\mathbf{A}}$ is one-to-one if and only if the only solution to $\mathbf{A}\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. We gather these facts in the following theorem.

www.RanaMaths.com

Theorem 8.11: Let $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ be a matrix mapping, where $\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n] \in M_{m\times n}$. The following statements are equivalent:

1. $\mathsf{T}_{\mathbf{A}}$ is one-to-one.
2. The rank of $\mathbf{A}$ is $r = \text{rank}(\mathbf{A}) = n$.
3. The columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent.

Example 8.12. Let $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^4 \to \mathbb{R}^3$ be the matrix mapping with matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -2 & 6 & 4 \\ -1 & 0 & -2 & -1 \\ 2 & -2 & 0 & 2 \end{bmatrix}.$$

Is $\mathsf{T}_{\mathbf{A}}$ one-to-one?

Solution. By Theorem 8.11, $\mathsf{T}_{\mathbf{A}}$ is one-to-one if and only if the columns of $\mathbf{A}$ are linearly independent. The columns of $\mathbf{A}$ lie in $\mathbb{R}^3$ and there are $n = 4$ columns. From Lecture 6, we know then that the columns are not linearly independent. Therefore, $\mathsf{T}_{\mathbf{A}}$ is not one-to-one. Alternatively, $\mathbf{A}$ will have rank at most $r = 3$ (why?). Therefore, the solution set to $\mathbf{Ax} = \mathbf{0}$ will have at least one parameter, and thus there exists infinitely many solutions to $\mathbf{Ax} = \mathbf{0}$. Intuitively, because $\mathbb{R}^4$ is "larger" than $\mathbb{R}^3$, the linear mapping $\mathsf{T}_{\mathbf{A}}$ will have to project $\mathbb{R}^4$ onto $\mathbb{R}^3$ and thus infinitely many vectors in $\mathbb{R}^4$ will be mapped to the same vector in $\mathbb{R}^3$. □

Example 8.13. Let $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^3$ be the matrix mapping with matrix

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

Is $\mathsf{T}_{\mathbf{A}}$ one-to-one?

Solution. By inspection, we see that the columns of $\mathbf{A}$ are linearly independent. Therefore, $\mathsf{T}_{\mathbf{A}}$ is one-to-one. Alternatively, one can compute that

$$\texttt{rref}(\mathbf{A}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Therefore, $r = \text{rank}(\mathbf{A}) = 2$, which is equal to the number columns of $\mathbf{A}$. □

www.RanaMaths.com

8.3 Standard Matrix of a Linear Mapping

We have shown that all matrix mappings $\mathsf{T}_{\mathbf{A}}$ are linear mappings. We now want to answer the reverse question: Are all linear mappings matrix mappings in disguise? If $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, then to show that T is in fact a matrix mapping we must show that there is some matrix $\mathbf{A} \in M_{m\times n}$ such that $\mathsf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. To that end, introduce the **standard unit vectors** $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ \cdots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Every $\mathbf{x} \in \mathbb{R}^n$ is in $\operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ because:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$$

With this notation we prove the following.

Theorem 8.14: Every linear mapping is a matrix mapping.

Proof. Let $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear mapping. Let

$$\mathbf{v}_1 = \mathsf{T}(\mathbf{e}_1), \mathbf{v}_2 = \mathsf{T}(\mathbf{e}_2), \ldots, \mathbf{v}_n = \mathsf{T}(\mathbf{e}_n).$$

The co-domain of T is $\mathbb{R}^m$, and thus $\mathbf{v}_i \in \mathbb{R}^m$. Now, for arbitrary $\mathbf{x} \in \mathbb{R}^n$ we can write

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n.$$

Then by linearity of T, we have

$$\begin{aligned} \mathsf{T}(\mathbf{x}) &= \mathsf{T}(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\ &= x_1\mathsf{T}(\mathbf{e}_1) + x_2\mathsf{T}(\mathbf{e}_2) + \cdots + x_n\mathsf{T}(\mathbf{e}_n) \\ &= x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n \\ &= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \mathbf{x}. \end{aligned}$$

Define the matrix $\mathbf{A} \in M_{m\times n}$ by $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$. Then our computation above shows that

$$\mathsf{T}(\mathbf{x}) = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{A}\mathbf{x}.$$

Therefore, T is a matrix mapping with the matrix $\mathbf{A} \in M_{m\times n}$. □

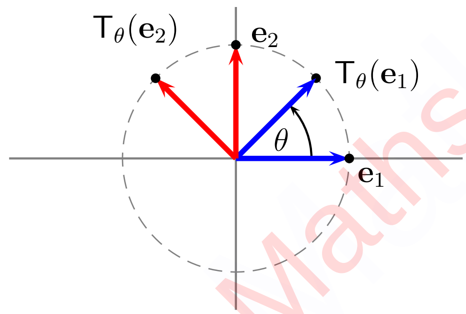
www.RanaMaths.com

If $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping, the matrix

$$\mathbf{A} = \begin{bmatrix} \mathsf{T}(\mathbf{e}_1) & \mathsf{T}(\mathbf{e}_2) & \cdots & \mathsf{T}(\mathbf{e}_n) \end{bmatrix}$$

is called the **standard matrix of** T. In words, the columns of $\mathbf{A}$ are the images of the standard unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ under T. The punchline is that if T is a linear mapping, then to derive properties of T we need only know the standard matrix $\mathbf{A}$ corresponding to T.

Example 8.15. Let $\mathsf{T} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear mapping that rotates every vector by an angle θ. Use the standard unit vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$ to write down the matrix $\mathbf{A} \in \mathbb{R}^{2\times 2}$ corresponding to T.



Solution. We have

$$\mathbf{A} = \begin{bmatrix} \mathsf{T}(\mathbf{e}_1) & \mathsf{T}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

□

Example 8.16. Let $\mathsf{T} : \mathbb{R}^3 \to \mathbb{R}^3$ be a dilation of factor $k = 2$. Find the standard matrix $\mathbf{A}$ of T.

Solution. The mapping is $\mathsf{T}(\mathbf{x}) = 2\mathbf{x}$. Then

$$\mathsf{T}(\mathbf{e}_1) = 2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad \mathsf{T}(\mathbf{e}_2) = 2\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \quad \mathsf{T}(\mathbf{e}_3) = 2\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Therefore,

$$\mathbf{A} = \begin{bmatrix} \mathsf{T}(\mathbf{e}_1) & \mathsf{T}(\mathbf{e}_2) & \mathsf{T}(\mathbf{e}_3) \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

is the standard matrix of T. □

After this lecture you should know the following:

www.RanaMaths.com

- the relationship between the range of a matrix mapping $\mathsf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ and the span of the columns of $\mathbf{A}$
- what it means for a mapping to be onto and one-to-one
- how to verify if a linear mapping is onto and one-to-one
- that all linear mappings are matrix mappings
- what the standard unit vectors are
- how to compute the standard matrix of a linear mapping

www.RanaMaths.com

www.RanaMaths.com

Lecture 9

Matrix Algebra

9.1 Sums of Matrices

We begin with the definition of matrix addition.

Definition 9.1: Given matrices

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

both of the same dimension $m \times n$, the sum $\mathbf{A} + \mathbf{B}$ is defined as

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

Next is the definition of scalar-matrix multiplication.

Definition 9.2: For a scalar α we define $\alpha\mathbf{A}$ by

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

www.RanaMaths.com

Example 9.3. Given **A** and **B** below, find $3\mathbf{A} - 2\mathbf{B}$.

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 5 \\ 0 & -3 & 9 \\ 4 & -6 & 7 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} 5 & 0 & -11 \\ 3 & -5 & 1 \\ -1 & -9 & 0 \end{bmatrix}$$

Solution. We compute:

$$3\mathbf{A} - 2\mathbf{B} = \begin{bmatrix} 3 & -6 & 15 \\ 0 & -9 & 27 \\ 12 & -18 & 21 \end{bmatrix} - \begin{bmatrix} 10 & 0 & -22 \\ 6 & -10 & 2 \\ -2 & -18 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -7 & -6 & 37 \\ -6 & 1 & 25 \\ 14 & 0 & 21 \end{bmatrix}$$

□

Below are some basic algebraic properties of matrix addition/scalar multiplication.

Theorem 9.4: Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be matrices of the same size and let α, β be scalars. Then

(a) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
(b) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
(c) $\mathbf{A} + \mathbf{0} = \mathbf{A}$
(d) $\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$
(e) $(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$
(f) $\alpha(\beta\mathbf{A}) = (\alpha\beta)\mathbf{A}$

9.2 Matrix Multiplication

Let $\mathsf{T}_\mathbf{B} : \mathbb{R}^p \to \mathbb{R}^n$ and let $\mathsf{T}_\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$ be linear mappings. If $\mathbf{x} \in \mathbb{R}^p$ then $\mathsf{T}_\mathbf{B}(\mathbf{x}) \in \mathbb{R}^n$ and thus we can apply $\mathsf{T}_\mathbf{A}$ to $\mathsf{T}_\mathbf{B}(\mathbf{x})$. The resulting vector $\mathsf{T}_\mathbf{A}(\mathsf{T}_\mathbf{B}(\mathbf{x}))$ is in $\mathbb{R}^m$. Hence, each $\mathbf{x} \in \mathbb{R}^p$ can be mapped to a point in $\mathbb{R}^m$, and because $\mathsf{T}_\mathbf{B}$ and $\mathsf{T}_\mathbf{A}$ are linear mappings the resulting mapping is also linear. This resulting mapping is called the **composition** of $\mathsf{T}_\mathbf{A}$ and $\mathsf{T}_\mathbf{B}$, and is usually denoted by $\mathsf{T}_\mathbf{A} \circ \mathsf{T}_\mathbf{B} : \mathbb{R}^p \to \mathbb{R}^m$ (see Figure 9.1). Hence,

$$(\mathsf{T}_\mathbf{A} \circ \mathsf{T}_\mathbf{B})(\mathbf{x}) = \mathsf{T}_\mathbf{A}(\mathsf{T}_\mathbf{B}(\mathbf{x})).$$

Because $(\mathsf{T}_\mathbf{A} \circ \mathsf{T}_\mathbf{B}) : \mathbb{R}^p \to \mathbb{R}^m$ is a linear mapping it has an associated standard matrix, which we denote for now by $\mathbf{C}$. From Lecture 8, to compute the standard matrix of any linear mapping, we must compute the images of the standard unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ under the linear mapping. Now, for any $\mathbf{x} \in \mathbb{R}^p$,

$$\mathsf{T}_\mathbf{A}(\mathsf{T}_\mathbf{B}(\mathbf{x})) = \mathsf{T}_\mathbf{A}(\mathbf{B}\mathbf{x}) = \mathbf{A}(\mathbf{B}\mathbf{x}).$$

Applying this to $\mathbf{x} = \mathbf{e}_i$ for all $i = 1, 2, \ldots, p$, we obtain the standard matrix of $\mathsf{T}_\mathbf{A} \circ \mathsf{T}_\mathbf{B}$:

$$\mathbf{C} = \begin{bmatrix} \mathbf{A}(\mathbf{B}\mathbf{e}_1) & \mathbf{A}(\mathbf{B}\mathbf{e}_2) & \cdots & \mathbf{A}(\mathbf{B}\mathbf{e}_p) \end{bmatrix}.$$

www.RanaMaths.com

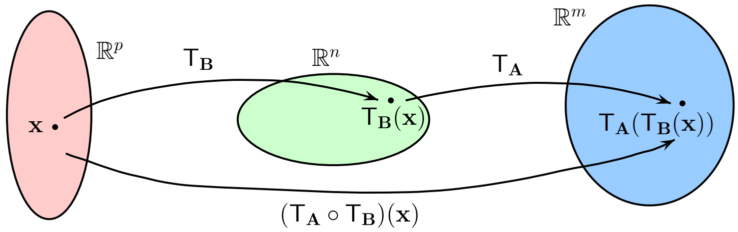


Figure 9.1: Illustration of the composition of two mappings.

Now $\mathbf{B}\mathbf{e}_1$ is

$$\mathbf{B}\mathbf{e}_1 = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix} \mathbf{e}_1 = \mathbf{b}_1.$$

And similarly $\mathbf{B}\mathbf{e}_i = \mathbf{b}_i$ for all $i = 1, 2, \ldots, p$. Therefore,

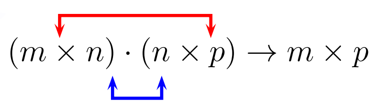
$$\mathbf{C} = \begin{bmatrix} \mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 & \cdots & \mathbf{A}\mathbf{b}_p \end{bmatrix}$$

is the standard matrix of $\mathsf{T}_{\mathbf{A}} \circ \mathsf{T}_{\mathbf{B}}$. This computation motivates the following definition.

Definition 9.5: For $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$, with $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \cdots & \mathbf{b}_p \end{bmatrix}$, we define the product $\mathbf{AB}$ by the formula

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 & \cdots & \mathbf{A}\mathbf{b}_p \end{bmatrix}.$$

The product $\mathbf{AB}$ is defined only when the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$. The following diagram is useful for remembering this:



From our definition of $\mathbf{AB}$, the standard matrix of the composite mapping $\mathsf{T}_{\mathbf{A}} \circ \mathsf{T}_{\mathbf{B}}$ is

$$\mathbf{C} = \mathbf{AB}.$$

In other words, composition of linear mappings corresponds to matrix multiplication.

Example 9.6. For $\mathbf{A}$ and $\mathbf{B}$ below compute $\mathbf{AB}$ and $\mathbf{BA}$.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & -3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -4 & 2 & 4 & -4 \\ -1 & -5 & -3 & 3 \\ -4 & -4 & -3 & -1 \end{bmatrix}$$

www.RanaMaths.com

Solution. First $\mathbf{AB} = [\mathbf{Ab}_1 \;\; \mathbf{Ab}_2 \;\; \mathbf{Ab}_3 \;\; \mathbf{Ab}_4]$:

$$\begin{aligned}
\mathbf{AB} &= \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & -3 \end{bmatrix} \begin{bmatrix} -4 & 2 & 4 & -4 \\ -1 & -5 & -3 & 3 \\ -4 & -4 & -3 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 2 \\ 7 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 \\ 7 & 9 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 & 4 \\ 7 & 9 & 10 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 0 & 4 & 4 \\ 7 & 9 & 10 & 2 \end{bmatrix}
\end{aligned}$$

On the other hand, $\mathbf{BA}$ is not defined! $\mathbf{B}$ has 4 columns and $\mathbf{A}$ has 2 rows. □

Example 9.7. For $\mathbf{A}$ and $\mathbf{B}$ below compute $\mathbf{AB}$ and $\mathbf{BA}$.

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

Solution. First $\mathbf{AB} = [\mathbf{Ab}_1 \;\; \mathbf{Ab}_2 \;\; \mathbf{Ab}_3]$:

$$\begin{aligned}
\mathbf{AB} &= \begin{bmatrix} -4 & 4 & 3 \\ 3 & -3 & -1 \\ -2 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 \\ -3 & 0 & -2 \\ -2 & 1 & -2 \end{bmatrix} \\
&= \begin{bmatrix} -14 \\ 8 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} -14 & 7 \\ 8 & -4 \\ 3 & 3 \end{bmatrix} \\
&= \begin{bmatrix} -14 & 7 & -14 \\ 8 & -4 & 8 \\ 3 & 3 & 0 \end{bmatrix}
\end{aligned}$$

Next $\mathbf{BA} = [\mathbf{Ba}_1 \;\; \mathbf{Ba}_2 \;\; \mathbf{Ba}_3]$:

$$\begin{aligned}
\mathbf{BA} &= \begin{bmatrix} -1 & -1 & 0 \\ -3 & 0 & -2 \\ -2 & 1 & -2 \end{bmatrix} \begin{bmatrix} -4 & 4 & 3 \\ 3 & -3 & -1 \\ -2 & -1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & & \\ 16 & & \\ 15 & & \end{bmatrix} \\
&= \begin{bmatrix} 1 & -1 & \\ 16 & -10 & \\ 15 & -9 & \end{bmatrix} \\
&= \begin{bmatrix} 1 & -1 & -2 \\ 16 & -10 & -11 \\ 15 & -9 & -9 \end{bmatrix}
\end{aligned}$$

On the other hand:

$$\mathbf{AB} = \begin{bmatrix} -14 & 7 & -14 \\ 8 & -4 & 8 \\ 3 & 3 & 0 \end{bmatrix}$$

Therefore, in general $\mathbf{AB} \neq \mathbf{BA}$, i.e., matrix multiplication is not commutative. □

An important matrix that arises frequently is the **identity matrix** $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ **of size** n:

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

You should verify that for any $\mathbf{A} \in \mathbb{R}^{n\times n}$ it holds that $\mathbf{AI}_n = \mathbf{I}_n\mathbf{A} = \mathbf{A}$. Below are some basic algebraic properties of matrix multiplication.

Theorem 9.8: Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be matrices, of appropriate dimensions, and let α be a scalar. Then

(1) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

(2) $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

(3) $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$

(4) $\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B})$

(5) $\mathbf{I}_n\mathbf{A} = \mathbf{AI}_n = \mathbf{A}$

If $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a square matrix, the kth power of $\mathbf{A}$ is

$$\mathbf{A}^k = \underbrace{\mathbf{AAA}\cdots\mathbf{A}}_{k \text{ times}}$$

www.RanaMaths.com

Example 9.9. Compute $\mathbf{A}^3$ if

$$\mathbf{A} = \begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix}.$$

Solution. Compute $\mathbf{A}^2$:

$$\mathbf{A}^2 = \begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 7 & -6 \\ -2 & 3 \end{bmatrix}$$

And then $\mathbf{A}^3$:

$$\mathbf{A}^3 = \mathbf{A}^2\mathbf{A} = \begin{bmatrix} 7 & -6 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -20 & 21 \\ 7 & -6 \end{bmatrix}$$

We could also do:

$$\mathbf{A}^3 = \mathbf{A}\mathbf{A}^2 = \begin{bmatrix} -2 & 3 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 7 & -6 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -20 & 21 \\ 7 & -6 \end{bmatrix}.$$

□

9.3 Matrix Transpose

We begin with the definition of the transpose of a matrix.

Definition 9.10: Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the **transpose** of $\mathbf{A}$ is the matrix $\mathbf{A}^T$ whose ith column is the ith row of $\mathbf{A}$.

If $\mathbf{A}$ is $m \times n$ then $\mathbf{A}^T$ is $n \times m$. For example, if

$$\mathbf{A} = \begin{bmatrix} 0 & -1 & 8 & -7 & -4 \\ -4 & 6 & -10 & -9 & 6 \\ 9 & 5 & -2 & -3 & 5 \\ -8 & 8 & 4 & 7 & 7 \end{bmatrix}$$

then

$$\mathbf{A}^T = \begin{bmatrix} 0 & -4 & 9 & -8 \\ -1 & 6 & 5 & 8 \\ 8 & -10 & -2 & 4 \\ -7 & -9 & -3 & 7 \\ -4 & 6 & 5 & 7 \end{bmatrix}.$$

Example 9.11. Compute $(\mathbf{AB})^T$ and $\mathbf{B}^T\mathbf{A}^T$ if

$$\mathbf{A} = \begin{bmatrix} -2 & 1 & 0 \\ 3 & -1 & -3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -2 & 1 & 2 \\ -1 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Solution. Compute $\mathbf{AB}$:

$$\mathbf{AB} = \begin{bmatrix} -2 & 1 & 0 \\ 3 & -1 & -3 \end{bmatrix} \begin{bmatrix} -2 & 1 & 2 \\ -1 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & -4 & -4 \\ -5 & 5 & 9 \end{bmatrix}$$

Next compute $\mathbf{B}^T\mathbf{A}^T$:

$$\mathbf{B}^T\mathbf{A}^T = \begin{bmatrix} -2 & -1 & 0 \\ 1 & -2 & 0 \\ 2 & 0 & -1 \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 1 & -1 \\ 0 & -3 \end{bmatrix} = \begin{bmatrix} 3 & -5 \\ -4 & 5 \\ -4 & 9 \end{bmatrix} = (\mathbf{AB})^T$$

□

The following theorem summarizes properties of the transpose.

Theorem 9.12: Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of appropriate sizes. The following hold:
(1) $(\mathbf{A}^T)^T = \mathbf{A}$
(2) $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
(3) $(\alpha\mathbf{A})^T = \alpha\mathbf{A}^T$
(4) $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

A consequence of property (4) is that

$$(\mathbf{A}_1\mathbf{A}_2\ldots\mathbf{A}_k)^T = \mathbf{A}_k^T\mathbf{A}_{k-1}^T\cdots\mathbf{A}_2^T\mathbf{A}_1^T$$

and as a special case

$$(\mathbf{A}^k)^T = (\mathbf{A}^T)^k.$$

Example 9.13. Let $\mathsf{T}: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear mapping that first contracts vectors by a factor of $k = 3$ and then rotates by an angle θ. What is the standard matrix $\mathbf{A}$ of T?

Solution. Let $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ denote the standard unit vectors in $\mathbb{R}^2$. From Lecture 8, the standard matrix of T is $\mathbf{A} = \begin{bmatrix}\mathsf{T}(\mathbf{e}_1) & \mathsf{T}(\mathbf{e}_2)\end{bmatrix}$. Recall that the standard matrix of a rotation by θ is

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Contracting $\mathbf{e}_1$ by a factor of $k = 3$ results in $(\frac{1}{3}, 0)$ and then rotation by θ results in

$$\begin{bmatrix} \frac{1}{3}\cos(\theta) \\ \frac{1}{3}\sin(\theta) \end{bmatrix} = \mathsf{T}(\mathbf{e}_1).$$

www.RanaMaths.com

Contracting $\mathbf{e}_2$ by a factor of $k = 3$ results in $(0, \frac{1}{3})$ and then rotation by θ results in

$$\begin{bmatrix} -\frac{1}{3}\sin(\theta) \\ \frac{1}{3}\cos(\theta) \end{bmatrix} = \mathsf{T}(\mathbf{e}_2).$$

Therefore,

$$\mathbf{A} = \begin{bmatrix} \mathsf{T}(\mathbf{e}_1) & \mathsf{T}(\mathbf{e}_2) \end{bmatrix} = \begin{bmatrix} \frac{1}{3}\cos(\theta) & -\frac{1}{3}\sin(\theta) \\ \frac{1}{3}\sin(\theta) & \frac{1}{3}\cos(\theta) \end{bmatrix}$$

On the other hand, the standard matrix corresponding to a contraction by a factor $k = \frac{1}{3}$ is

$$\begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

Therefore,

$$\underbrace{\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}}_{\text{rotation}} \underbrace{\begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}}_{\text{contraction}} = \begin{bmatrix} \frac{1}{3}\cos(\theta) & -\frac{1}{3}\sin(\theta) \\ \frac{1}{3}\sin(\theta) & \frac{1}{3}\cos(\theta) \end{bmatrix} = \mathbf{A}$$

□

After this lecture you should know the following:

- know how to add and multiply matrices
- that matrix multiplication corresponds to composition of linear mappings
- the algebraic properties of matrix multiplication (Theorem 9.8)
- how to compute the transpose of a matrix
- the properties of matrix transposition (Theorem 9.12)

www.RanaMaths.com

Lecture 10

Invertible Matrices

10.1 Inverse of a Matrix

The inverse of a **square** matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ generalizes the notion of the reciprocal of a non-zero number $a \in \mathbb{R}$. Formally speaking, the inverse of a non-zero number $a \in \mathbb{R}$ is the unique number $c \in \mathbb{R}$ such that $ac = ca = 1$. The inverse of $a \neq 0$, usually denoted by $a^{-1} = \frac{1}{a}$, can be used to solve the equation $ax = b$:

$$ax = b \;\Rightarrow\; a^{-1}ax = a^{-1}b \;\Rightarrow\; x = a^{-1}b.$$

This motivates the following definition.

Definition 10.1: A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is called **invertible** if there exists a matrix $\mathbf{C} \in \mathbb{R}^{n\times n}$ such that $\mathbf{AC} = \mathbf{I}_n$ and $\mathbf{CA} = \mathbf{I}_n$.

If $\mathbf{A}$ is invertible then can it have more than one inverse? Suppose that there exists $\mathbf{C}_1, \mathbf{C}_2$ such that $\mathbf{AC}_i = \mathbf{C}_i\mathbf{A} = \mathbf{I}_n$. Then

$$\mathbf{C}_2 = \mathbf{C}_2(\mathbf{AC}_1) = (\mathbf{C}_2\mathbf{A})\mathbf{C}_1 = \mathbf{I}_n\mathbf{C}_1 = \mathbf{C}_1.$$

Thus, if $\mathbf{A}$ is invertible, it can have only one inverse. This motivates the following definition.

Definition 10.2: If $\mathbf{A}$ is invertible then we denote **the** inverse of $\mathbf{A}$ by $\mathbf{A}^{-1}$. Thus, $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$.

Example 10.3. Given $\mathbf{A}$ and $\mathbf{C}$ below, show that $\mathbf{C}$ is the inverse of $\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix}$$

www.RanaMaths.com

Solution. Compute $\mathbf{AC}$:

$$\mathbf{AC} = \begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix} \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Compute $\mathbf{CA}$:

$$\mathbf{CA} = \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Therefore, by definition $\mathbf{C} = \mathbf{A}^{-1}$. □

Theorem 10.4: Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and suppose that $\mathbf{A}$ is invertible. Then for any $\mathbf{b} \in \mathbb{R}^n$ the matrix equation $\mathbf{Ax} = \mathbf{b}$ has a unique solution given by $\mathbf{A}^{-1}\mathbf{b}$.

Proof: Let $\mathbf{b} \in \mathbb{R}^n$ be arbitrary. Then multiplying the equation $\mathbf{Ax} = \mathbf{b}$ by $\mathbf{A}^{-1}$ from the left we obtain that

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{Ax} &= \mathbf{A}^{-1}\mathbf{b} \\ \Rightarrow \mathbf{I}_n\mathbf{x} &= \mathbf{A}^{-1}\mathbf{b} \\ \Rightarrow \mathbf{x} &= \mathbf{A}^{-1}\mathbf{b}. \end{aligned}$$

Therefore, with $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ we have that

$$\mathbf{Ax} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = \mathbf{AA}^{-1}\mathbf{b} = \mathbf{I}_n\mathbf{b} = \mathbf{b}$$

and thus $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ is a solution. If $\tilde{\mathbf{x}}$ is another solution of the equation, that is, $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}$, then multiplying both sides by $\mathbf{A}^{-1}$ we obtain that $\tilde{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{b}$. Thus, $\mathbf{x} = \tilde{\mathbf{x}}$. □

Example 10.5. Use the result of **Example 10.3.** to solve the linear system $\mathbf{Ax} = \mathbf{b}$ if

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix}.$$

Solution. We showed in Example 10.3 that

$$\mathbf{A}^{-1} = \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix}.$$

Therefore, the unique solution to the linear system $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{A}^{-1}\mathbf{b} = \begin{bmatrix} -14 & -3 & -6 \\ -5 & -1 & -2 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

www.RanaMaths.com

Verify:

$$\begin{bmatrix} 1 & -3 & 0 \\ -1 & 2 & -2 \\ -2 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix}$$

□

The following theorem summarizes the relationship between the matrix inverse and matrix multiplication and matrix transpose.

Theorem 10.6: Let $\mathbf{A}$ and $\mathbf{B}$ be invertible matrices. Then:

(1) The matrix $\mathbf{A}^{-1}$ is invertible and its inverse is $\mathbf{A}$:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}.$$

(2) The matrix $\mathbf{AB}$ is invertible and its inverse is $\mathbf{B}^{-1}\mathbf{A}^{-1}$:

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

(3) The matrix $\mathbf{A}^T$ is invertible and its inverse is $(\mathbf{A}^{-1})^T$:

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T.$$

Proof: To prove **(2)** we compute

$$(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AI}_n\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}_n.$$

To prove **(3)** we compute

$$\mathbf{A}^T(\mathbf{A}^{-1})^T = (\mathbf{A}^{-1}\mathbf{A})^T = \mathbf{I}_n^T = \mathbf{I}_n.$$

□

10.2 Computing the Inverse of a Matrix

If $\mathbf{A} \in M_{n\times n}$ is invertible, how do we find $\mathbf{A}^{-1}$? Let $\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \end{bmatrix}$ and we will find expressions for $\mathbf{c}_i$. First note that $\mathbf{AA}^{-1} = \begin{bmatrix} \mathbf{Ac}_1 & \mathbf{Ac}_2 & \cdots & \mathbf{Ac}_n \end{bmatrix}$. On the other hand, we also have $\mathbf{AA}^{-1} = \mathbf{I}_n = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}$. Therefore, we want to find $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ such that

$$\underbrace{\begin{bmatrix} \mathbf{Ac}_1 & \mathbf{Ac}_2 & \cdots & \mathbf{Ac}_n \end{bmatrix}}_{\mathbf{AA}^{-1}} = \underbrace{\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}}_{\mathbf{I}_n}.$$

To find $\mathbf{c}_i$ we therefore need to solve the linear system $\mathbf{Ax} = \mathbf{e}_i$. Here the image vector "$\mathbf{b}$" is $\mathbf{e}_i$. To find $\mathbf{c}_1$ we form the augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{e}_1 \end{bmatrix}$ and find its RREF:

$$\begin{bmatrix} \mathbf{A} & \mathbf{e}_1 \end{bmatrix} \sim \begin{bmatrix} \mathbf{I}_n & \mathbf{c}_1 \end{bmatrix}.$$

www.RanaMaths.com

We will need to do this for each $\mathbf{c}_2, \ldots, \mathbf{c}_n$ so we might as well form the combined augmented matrix $\begin{bmatrix}\mathbf{A} & \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n\end{bmatrix}$ and find the RREF all at once:

$$\begin{bmatrix}\mathbf{A} & \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n\end{bmatrix} \sim \begin{bmatrix}\mathbf{I}_n & \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n\end{bmatrix}.$$

In summary, to determine if $\mathbf{A}^{-1}$ exists and to simultaneously compute it, we compute the RREF of the augmented matrix

$$\begin{bmatrix}\mathbf{A} & \mathbf{I}_n\end{bmatrix},$$

that is, $\mathbf{A}$ augmented with the $n \times n$ identity matrix. If the RREF of $\mathbf{A}$ is $\mathbf{I}_n$, that is

$$\begin{bmatrix}\mathbf{A} & \mathbf{I}_n\end{bmatrix} \sim \begin{bmatrix}\mathbf{I}_n & \mathbf{c}_1 & \mathbf{c}_2 \cdots & \mathbf{c}_n\end{bmatrix}$$

then

$$\mathbf{A}^{-1} = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 \cdots & \mathbf{c}_n\end{bmatrix}.$$

If the RREF of $\mathbf{A}$ is not $\mathbf{I}_n$ then $\mathbf{A}$ is not invertible.

Example 10.7. Find the inverse of $\mathbf{A} = \begin{bmatrix}1 & 3\\ -1 & -2\end{bmatrix}$ if it exists.

Solution. Form the augmented matrix $\begin{bmatrix}\mathbf{A} & \mathbf{I}_2\end{bmatrix}$ and row reduce:

$$\begin{bmatrix}\mathbf{A} & \mathbf{I}_2\end{bmatrix} = \begin{bmatrix}1 & 3 & 1 & 0\\ -1 & -2 & 0 & 1\end{bmatrix}$$

Add rows R_1 and R_2:

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

Perform the operation $\xrightarrow{-3R_2+R_1}$:

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

Thus, $\texttt{rref}(\mathbf{A}) = \mathbf{I}_2$, and therefore $\mathbf{A}$ is invertible. The inverse is

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

Verify:

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix}1 & 3\\ -1 & -2\end{bmatrix}\begin{bmatrix}-2 & -3\\ 1 & 1\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

□

Example 10.8. Find the inverse of $\mathbf{A} = \begin{bmatrix}1 & 0 & 3\\ 1 & 1 & 0\\ -2 & 0 & -7\end{bmatrix}$ if it exists.

Solution. Form the augmented matrix $\begin{bmatrix}\mathbf{A} & \mathbf{I}_3\end{bmatrix}$ and row reduce:

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

$-R_3$:

$$\begin{bmatrix} 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -1 & 1 & 0 \\ 0 & 0 & -1 & 2 & 0 & 1 \end{bmatrix} \xrightarrow{-R_3} \begin{bmatrix} 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -1 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 & -1 \end{bmatrix}$$

$3R_3 + R_2$ and $-3R_3 + R_1$:

$$\begin{bmatrix} 1 & 0 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -1 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 & -1 \end{bmatrix} \xrightarrow{3R_3+R_2,\ -3R_3+R_1} \begin{bmatrix} 1 & 0 & 0 & 7 & 0 & 3 \\ 0 & 1 & 0 & -7 & 1 & -3 \\ 0 & 0 & 1 & -2 & 0 & -1 \end{bmatrix}$$

Therefore, $\texttt{rref}(\mathbf{A}) = \mathbf{I}_3$, and therefore $\mathbf{A}$ is invertible. The inverse is

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

Verify:

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

□

Example 10.9. Find the inverse of $\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & -2 \\ -2 & 0 & -2 \end{bmatrix}$ if it exists.

Solution. Form the augmented matrix $\begin{bmatrix}\mathbf{A} & \mathbf{I}_3\end{bmatrix}$ and row reduce:

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

We need not go further since the $\texttt{rref}(\mathbf{A})$ is not $\mathbf{I}_3$ ($\text{rank}(\mathbf{A}) = 2$). Therefore, $\mathbf{A}$ is not invertible. □

10.3 Invertible Linear Mappings

Let $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^n$ be a matrix mapping with standard matrix $\mathbf{A}$ and suppose that $\mathbf{A}$ is invertible. Let $\mathsf{T}_{\mathbf{A}^{-1}} : \mathbb{R}^n \to \mathbb{R}^n$ be the matrix mapping with standard matrix $\mathbf{A}^{-1}$. Then the standard matrix of the composite mapping $\mathsf{T}_{\mathbf{A}^{-1}} \circ \mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^n$ is

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n.$$

www.RanaMaths.com

Therefore, $(\mathsf{T}_{\mathbf{A}^{-1}} \circ \mathsf{T}_{\mathbf{A}})(\mathbf{x}) = \mathbf{I}_n\mathbf{x} = \mathbf{x}$. Let's unravel $(\mathsf{T}_{\mathbf{A}^{-1}} \circ \mathsf{T}_{\mathbf{A}})(\mathbf{x})$ to see this:

$$(\mathsf{T}_{\mathbf{A}^{-1}} \circ \mathsf{T}_{\mathbf{A}})(\mathbf{x}) = \mathsf{T}_{\mathbf{A}^{-1}}(\mathsf{T}_{\mathbf{A}}(\mathbf{x})) = \mathsf{T}_{\mathbf{A}^{-1}}(\mathbf{A}\mathbf{x}) = \mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{x}.$$

Similarly, the standard matrix of $(\mathsf{T}_{\mathbf{A}} \circ \mathsf{T}_{\mathbf{A}^{-1}})$ is also $\mathbf{I}_n$. Intuitively, the linear mapping $\mathsf{T}_{\mathbf{A}^{-1}}$ *undoes* what $\mathsf{T}_{\mathbf{A}}$ does, and conversely. Moreover, since $\mathbf{A}\mathbf{x} = \mathbf{b}$ always has a solution, $\mathsf{T}_{\mathbf{A}}$ is onto. And, because the solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ is unique, $\mathsf{T}_{\mathbf{A}}$ is one-to-one.

The following theorem summarizes equivalent conditions for matrix invertibility.

Theorem 10.10: Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. The following statements are equivalent:

(a) $\mathbf{A}$ is invertible.
(b) $\mathbf{A}$ is row equivalent to $\mathbf{I}_n$, that is, $\texttt{rref}(\mathbf{A}) = \mathbf{I}_n$.
(c) The equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(d) The linear transformation $\mathsf{T}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is one-to-one.
(e) The linear transformation $\mathsf{T}_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is onto.
(f) The matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ is always solvable.
(g) The columns of $\mathbf{A}$ span $\mathbb{R}^n$.
(h) The columns of $\mathbf{A}$ are linearly independent.
(i) $\mathbf{A}^T$ is invertible.

Proof: This is a summary of all the statements we have proved about matrices and matrix mappings specialized to the case of square matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$. Note that for non-square matrices, one-to-one does not imply ontoness, and conversely.

Example 10.11. Without doing any arithmetic, write down the inverse of the dilation matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}.$$

Example 10.12. Without doing any arithmetic, write down the inverse of the rotation matrix

$$\mathbf{A} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

After this lecture you should know the following:

- how to compute the inverse of a matrix
- properties of matrix inversion and matrix multiplication
- relate invertibility of a matrix with properties of the associated linear mapping (1-1, onto)
- the characterizations of invertible matrices Theorem 10.10

www.RanaMaths.com

Lecture 11

Determinants

11.1 Determinants of 2×2 and 3×3 Matrices

Consider a general 2×2 linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2. \end{aligned}$$

Using elementary row operations, it can be shown that the solution is

$$x_1 = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11}a_{22} - a_{12}a_{21}}, \quad x_2 = \frac{b_2 a_{11} - b_1 a_{21}}{a_{11}a_{22} - a_{12}a_{21}},$$

provided that $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Notice the denominator is the same in both expressions. The number $a_{11}a_{22} - a_{12}a_{21}$ then completely characterizes when a 2×2 linear system has a unique solution. This motivates the following definition.

Definition 11.1: Given a 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

we define the **determinant** of **A** as

$$\det \mathbf{A} = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

An alternative notation for $\det \mathbf{A}$ is using vertical bars:

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

www.RanaMaths.com

Example 11.2. Compute the determinant of $\mathbf{A}$.

(i) $\mathbf{A} = \begin{bmatrix} 3 & -1 \\ 8 & 2 \end{bmatrix}$

(ii) $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ -6 & -2 \end{bmatrix}$

(iii) $\mathbf{A} = \begin{bmatrix} -110 & 0 \\ 568 & 0 \end{bmatrix}$

Solution. For (i):

$$\det(\mathbf{A}) = \begin{vmatrix} 3 & -1 \\ 8 & 2 \end{vmatrix} = (3)(2) - (8)(-1) = 14$$

For (ii):

$$\det(\mathbf{A}) = \begin{vmatrix} 3 & 1 \\ -6 & -2 \end{vmatrix} = (3)(-2) - (-6)(1) = 0$$

For (iii):

$$\det(\mathbf{A}) = \begin{vmatrix} -110 & 0 \\ 568 & 0 \end{vmatrix} = (-110)(0) - (568)(0) = 0$$

□

As in the 2×2 case, the solution of a 3×3 linear system $\mathbf{Ax} = \mathbf{b}$ can be shown to be

$$x_1 = \frac{\text{Numerator}_1}{D}, \; x_2 = \frac{\text{Numerator}_2}{D}, \; x_3 = \frac{\text{Numerator}_3}{D}$$

where

$$D = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).$$

Notice that the terms of D in the parenthesis are determinants of 2×2 submatrices of $\mathbf{A}$:

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

Let

$$\mathbf{A}_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{A}_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}, \quad \text{and } \mathbf{A}_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

Then we can write

$$D = a_{11}\det(\mathbf{A}_{11}) - a_{12}\det(\mathbf{A}_{12}) + a_{13}\det(\mathbf{A}_{13}).$$

The matrix $\mathbf{A}_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$ is obtained from $\mathbf{A}$ by deleting the 1st row and the 1st column:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & \mathbf{a_{22}} & \mathbf{a_{23}} \\ a_{31} & \mathbf{a_{32}} & \mathbf{a_{33}} \end{bmatrix} \longrightarrow \mathbf{A}_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}.$$

Similarly, the matrix $\mathbf{A}_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}$ is obtained from $\mathbf{A}$ by deleting the 1st row and the 2nd column:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \mathbf{a_{21}} & a_{22} & \mathbf{a_{23}} \\ \mathbf{a_{31}} & a_{32} & \mathbf{a_{33}} \end{bmatrix} \longrightarrow \mathbf{A}_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}.$$

Finally, the matrix $\mathbf{A}_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$ is obtained from $\mathbf{A}$ by deleting the 1st row and the 3rd column:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \mathbf{a_{21}} & \mathbf{a_{22}} & a_{23} \\ \mathbf{a_{31}} & \mathbf{a_{32}} & a_{33} \end{bmatrix} \longrightarrow \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

Notice also that the sign in front of the coefficients a_{11}, a_{12}, and a_{13}, alternate. This motivates the following definition.

Definition 11.3: Let $\mathbf{A}$ be a 3×3 matrix. Let $\mathbf{A}_{jk}$ be the 2×2 matrix obtained from $\mathbf{A}$ by deleting the jth row and kth column. Define the **cofactor** of a_{jk} to be the number $C_{jk} = (-1)^{j+k} \det \mathbf{A}_{jk}$. Define the **determinant** of $\mathbf{A}$ to be

$$\det \mathbf{A} = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}.$$

This definition of the determinant is called the **expansion of the determinant along the first row**. In the cofactor $C_{jk} = (-1)^{j+k} \det \mathbf{A}_{jk}$, the expression $(-1)^{j+k}$ will evaluate to either 1 or -1, depending on whether $j + k$ is even or odd. For example, the cofactor of a_{12} is

$$C_{12} = (-1)^{1+2} \det \mathbf{A}_{12} = -\det \mathbf{A}_{12}$$

and the cofactor of a_{13} is

$$C_{13} = (-1)^{1+3} \det \mathbf{A}_{13} = \det \mathbf{A}_{13}.$$

We can also compute the cofactor of the other entries of $\mathbf{A}$ in the obvious way. For example, the cofactor of a_{23} is

$$C_{23} = (-1)^{2+3} \det \mathbf{A}_{23} = -\det \mathbf{A}_{23}.$$

A helpful way to remember the sign $(-1)^{j+k}$ of a cofactor is to use the matrix

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}.$$

This works not just for 3×3 matrices but for any square $n \times n$ matrix.

Example 11.4. Compute the determinant of the matrix

$$\mathbf{A} = \begin{bmatrix} 4 & -2 & 3 \\ 2 & 3 & 5 \\ 1 & 0 & 6 \end{bmatrix}$$

Solution. From the definition of the determinant

$$\begin{aligned}
\det \mathbf{A} &= a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} \\
&= (4)\det \mathbf{A}_{11} - (-2)\det \mathbf{A}_{12} + (3)\det \mathbf{A}_{13} \\
&= 4\begin{vmatrix} 3 & 5 \\ 0 & 6 \end{vmatrix} + 2\begin{vmatrix} 2 & 5 \\ 1 & 6 \end{vmatrix} + 3\begin{vmatrix} 2 & 3 \\ 1 & 0 \end{vmatrix} \\
&= 4(3 \cdot 6 - 5 \cdot 0) + 2(2 \cdot 6 - 1 \cdot 5) + 3(2 \cdot 0 - 1 \cdot 3) \\
&= 72 + 14 - 9 \\
&= 77
\end{aligned}$$

□

We can compute the determinant of a matrix $\mathbf{A}$ by expanding along any row or column. For example, the expansion of the determinant for the matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

along the 3rd row is

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

And along the 2nd column:

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

The punchline is that any way you choose to expand (row or column) you will get the same answer. If a particular row or column contains zeros, say entry a_{jk}, then the computation of the determinant is simplified if you expand along either row j or column k because $a_{jk}C_{jk} = 0$ and we need not compute C_{jk}.

Example 11.5. Compute the determinant of the matrix

$$\mathbf{A} = \begin{bmatrix} 4 & -2 & 3 \\ 2 & 3 & 5 \\ 1 & 0 & 6 \end{bmatrix}$$

Solution. In Example 11.4, we computed $\det(\mathbf{A}) = 77$ by expanding along the 1st row.

www.RanaMaths.com

Notice that $a_{32} = 0$. Expanding along the 3rd row:

$$\begin{aligned}\det \mathbf{A} &= (1)\det \mathbf{A}_{31} - (0)\det \mathbf{A}_{32} + (6)\det \mathbf{A}_{33} \\ &= \begin{vmatrix} -2 & 3 \\ 3 & 5 \end{vmatrix} + 6\begin{vmatrix} 4 & -2 \\ 2 & 3 \end{vmatrix} \\ &= 1(-2\cdot 5 - 3\cdot 3) + 6(4\cdot 3 - (-2)\cdot 2) \\ &= -19 + 96 \\ &= 77\end{aligned}$$

□

11.2 Determinants of $n \times n$ Matrices

Using the 3×3 case as a guide, we define the determinant of a general $n \times n$ matrix as follows.

Definition 11.6: Let $\mathbf{A}$ be a $n \times n$ matrix. Let $\mathbf{A}_{jk}$ be the $(n-1)\times(n-1)$ matrix obtained from $\mathbf{A}$ by deleting the jth row and kth column, and let $C_{jk} = (-1)^{j+k}\det \mathbf{A}_{jk}$ be the (j,k)**-cofactor** of $\mathbf{A}$. The determinant of $\mathbf{A}$ is defined to be

$$\det \mathbf{A} = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}.$$

The next theorem tells us that we can compute the determinant by expanding along any row or column.

Theorem 11.7: Let $\mathbf{A}$ be a $n \times n$ matrix. Then $\det \mathbf{A}$ may be obtained by a cofactor expansion along any row or any column of $\mathbf{A}$:

$$\det \mathbf{A} = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn}.$$

We obtain two immediate corollaries.

Corollary 11.8: If $\mathbf{A}$ has a row or column containing all zeros then $\det \mathbf{A} = 0$.

Proof. If the jth row contains all zeros then $a_{j1} = a_{j2} = \cdots = a_{jn} = 0$:

$$\det \mathbf{A} = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn} = 0.$$

www.RanaMaths.com

Corollary 11.9: For any square matrix $\mathbf{A}$ it holds that $\det \mathbf{A} = \det \mathbf{A}^T$.

Sketch of the proof. Expanding along the jth row of $\mathbf{A}$ is equivalent to expanding along the jth column of $\mathbf{A}^T$.

Example 11.10. Compute the determinant of

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

Solution. The third row contains two zeros, so expand along this row:

$$\begin{aligned} \det \mathbf{A} &= 0 \det \mathbf{A}_{31} - 0 \det \mathbf{A}_{32} + 2 \det \mathbf{A}_{33} - \det \mathbf{A}_{34} \\ &= 2 \begin{vmatrix} 1 & 3 & -2 \\ 1 & 2 & -1 \\ -1 & -3 & 0 \end{vmatrix} - \begin{vmatrix} 1 & 3 & 0 \\ 1 & 2 & -2 \\ -1 & -3 & 1 \end{vmatrix} \\ &= 2 \left(1 \begin{vmatrix} 2 & -1 \\ -3 & 0 \end{vmatrix} - 3 \begin{vmatrix} 1 & -1 \\ -1 & 0 \end{vmatrix} - 2 \begin{vmatrix} 1 & 2 \\ -1 & -3 \end{vmatrix} \right) \\ &\quad - \left(1 \begin{vmatrix} 2 & -2 \\ -3 & 1 \end{vmatrix} - 3 \begin{vmatrix} 1 & -2 \\ -1 & 1 \end{vmatrix} \right) \\ &= 2((0-3) - 3(0-1) - 2(-3+2)) - ((2-6) - 3(1-2)) \\ &= 5 \end{aligned}$$

□

Example 11.11. Compute the determinant of

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

www.RanaMaths.com

Solution. Expanding along the second row:

$$\det \mathbf{A} = -\det \mathbf{A}_{21} + 2\det \mathbf{A}_{22} - (-2)\det \mathbf{A}_{23} - 1\det \mathbf{A}_{24}$$

$$= -\begin{vmatrix} 3 & 0 & -2 \\ 0 & 2 & 1 \\ -3 & 1 & 0 \end{vmatrix} + 2\begin{vmatrix} 1 & 0 & -2 \\ 0 & 2 & 1 \\ -1 & 1 & 0 \end{vmatrix}$$

$$+ 2\begin{vmatrix} 1 & 3 & -2 \\ 0 & 0 & 1 \\ -1 & -3 & 0 \end{vmatrix} - \begin{vmatrix} 1 & 3 & 0 \\ 0 & 0 & 2 \\ -1 & -3 & 1 \end{vmatrix}$$

$$= -1(-3-12) + 2(-1-4) + 2(0) - (0)$$
$$= 5$$

□

11.3 Triangular Matrices

Below we introduce a class of matrices for which the determinant computation is trivial.

Definition 11.12: A square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is called **upper triangular** if $a_{jk} = 0$ whenever $j > k$. In other words, all the entries of $\mathbf{A}$ below the diagonal entries a_{ii} are zero. It is called **lower triangular** if $a_{jk} = 0$ whenever $j < k$.

For example, a 4×4 upper triangular matrix takes the form

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

Expanding along the first column, we compute

$$\det \mathbf{A} = a_{11}\begin{vmatrix} a_{22} & a_{23} & a_{24} \\ 0 & a_{33} & a_{34} \\ 0 & 0 & a_{44} \end{vmatrix} = a_{11}\left(a_{22}\begin{vmatrix} a_{33} & a_{34} \\ 0 & a_{44} \end{vmatrix}\right) = a_{11}a_{22}a_{33}a_{44}.$$

The general $n \times n$ case is similar and is summarized in the following theorem.

Theorem 11.13: The determinant of a triangular matrix is the product of its diagonal entries.

After this lecture you should know the following:

- how to compute the determinant of any sized matrix
- that the determinant of $\mathbf{A}$ is equal to the determinant of $\mathbf{A}^T$
- the determinant of a triangular matrix is the product of its diagonal entries

www.RanaMaths.com

www.RanaMaths.com

Lecture 12

Properties of the Determinant

12.1 ERO and Determinants

Recall that for a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ we defined

$$\det \mathbf{A} = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn}$$

where the number $C_{jk} = (-1)^{j+k} \det \mathbf{A}_{jk}$ is called the (j,k)-cofactor of $\mathbf{A}$ and

$$\mathbf{a}_j = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{jn} \end{bmatrix}$$

denotes the jth row of $\mathbf{A}$. Notice that

$$\det \mathbf{A} = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{jn} \end{bmatrix} \begin{bmatrix} C_{j1} \\ C_{j2} \\ \vdots \\ C_{jn} \end{bmatrix}.$$

If we let $\mathbf{c}_j = \begin{bmatrix} C_{j1} & C_{j2} & \cdots & C_{jn} \end{bmatrix}$ then

$$\det \mathbf{A} = \mathbf{a}_j \cdot \mathbf{c}_j^T.$$

In this lecture, we will establish properties of the determinant under elementary row operations and some consequences. The following theorem describes how the determinant behaves under elementary row operations of Type 1.

Theorem 12.1: Suppose that $\mathbf{A} \in \mathbb{R}^{n\times n}$ and let $\mathbf{B}$ be the matrix obtained by interchanging two rows of $\mathbf{A}$. Then $\det \mathbf{B} = -\det \mathbf{A}$.

Proof. Consider the 2×2 case. Let $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and let $\mathbf{B} = \begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix}$. Then

$$\det \mathbf{B} = a_{12}a_{21} - a_{11}a_{22} = -(a_{11}a_{22} - a_{12}a_{21}) = -\det \mathbf{A}.$$

The general case is proved by induction. □

This theorem leads to the following corollary.

www.RanaMaths.com

Corollary 12.2: If $\mathbf{A} \in \mathbb{R}^{n\times n}$ has two rows (or two columns) that are equal then $\det(\mathbf{A}) = 0$.

Proof. Suppose that $\mathbf{A}$ has rows j and k that are equal. Let $\mathbf{B}$ be the matrix obtained by interchanging rows j and k. Then by the previous theorem $\det\mathbf{B} = -\det\mathbf{A}$. But clearly $\mathbf{B} = \mathbf{A}$, and therefore $\det\mathbf{B} = \det\mathbf{A}$. Therefore, $\det(\mathbf{A}) = -\det(\mathbf{A})$ and thus $\det\mathbf{A} = 0$.
□ □

Now we consider how the determinant behaves under elementary row operations of Type 2.

Theorem 12.3: Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and let $\mathbf{B}$ be the matrix obtained by multiplying a row of $\mathbf{A}$ by β. Then $\det\mathbf{B} = \beta\det\mathbf{A}$.

Proof. Suppose that $\mathbf{B}$ is obtained from $\mathbf{A}$ by multiplying the jth row by β. The rows of $\mathbf{A}$ and $\mathbf{B}$ different from j are equal, and therefore

$$\mathbf{B}_{jk} = \mathbf{A}_{jk}, \quad \text{for } k = 1, 2, \ldots, n.$$

In particular, the (j,k) cofactors of $\mathbf{A}$ and $\mathbf{B}$ are equal. The jth row of $\mathbf{B}$ is $\beta\mathbf{a}_j$. Then, expanding $\det\mathbf{B}$ along the jth row:

$$\begin{aligned}\det\mathbf{B} &= (\beta\mathbf{a}_j)\cdot\mathbf{c}_j^T \\ &= \beta(\mathbf{a}_j\cdot\mathbf{c}_j^T) \\ &= \beta\det\mathbf{A}.\end{aligned}$$

□

Lastly we consider Type 3 elementary row operations.

Theorem 12.4: Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and let $\mathbf{B}$ be the matrix obtained from $\mathbf{A}$ by adding β times the kth row to the jth row. Then $\det\mathbf{B} = \det\mathbf{A}$.

Proof. For any matrix $\mathbf{A}$ and any row vector $\mathbf{r} = [r_1\ r_2\ \cdots\ r_n]$ the expression

$$\mathbf{r}\cdot\mathbf{c}_j^T = r_1C_{j1} + r_2C_{j2} + \cdots + r_nC_{jn}$$

is the determinant of the matrix obtained from $\mathbf{A}$ by replacing the jth row with the row $\mathbf{r}$. Therefore, if $k \neq j$ then

$$\mathbf{a}_k\cdot\mathbf{c}_j^T = 0$$

www.RanaMaths.com

since then rows k and j are equal. The jth row of $\mathbf{B}$ is $\mathbf{b}_j = \mathbf{a}_j + \beta\mathbf{a}_k$. Therefore, expanding $\det \mathbf{B}$ along the jth row:

$$\begin{aligned}\det \mathbf{B} &= (\mathbf{a}_j + \beta\mathbf{a}_k) \cdot \mathbf{c}_j^T \\ &= \mathbf{a}_j \cdot \mathbf{c}_j^T + \beta\left(\mathbf{a}_k \cdot \mathbf{c}_j^T\right) \\ &= \det \mathbf{A}.\end{aligned}$$

□

Example 12.5. Suppose that $\mathbf{A}$ is a 4×4 matrix and suppose that $\det \mathbf{A} = 11$. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by interchanging rows 2 and 4, what is $\det \mathbf{B}$?

Solution. Interchanging (or swapping) rows changes the sign of the determinant. Therefore,

$$\det \mathbf{B} = -11.$$

□

Example 12.6. Suppose that $\mathbf{A}$ is a 4×4 matrix and suppose that $\det \mathbf{A} = 11$. Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ denote the rows of $\mathbf{A}$. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by replacing row $\mathbf{a}_3$ by $3\mathbf{a}_1 + \mathbf{a}_3$, what is $\det \mathbf{B}$?

Solution. This is a Type 3 elementary row operation, which preserves the value of the determinant. Therefore,

$$\det \mathbf{B} = 11.$$

□

Example 12.7. Suppose that $\mathbf{A}$ is a 4×4 matrix and suppose that $\det \mathbf{A} = 11$. Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ denote the rows of $\mathbf{A}$. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by replacing row $\mathbf{a}_3$ by $3\mathbf{a}_1 + 7\mathbf{a}_3$, what is $\det \mathbf{B}$?

Solution. This is not quite a Type 3 elementary row operation because $\mathbf{a}_3$ is multiplied by 7. The third row of $\mathbf{B}$ is $\mathbf{b}_3 = 3\mathbf{a}_1 + 7\mathbf{a}_3$. Therefore, expanding $\det \mathbf{B}$ along the third row

$$\begin{aligned}\det \mathbf{B} &= (3\mathbf{a}_1 + 7\mathbf{a}_3) \cdot \mathbf{c}_3^T \\ &= 3\mathbf{a}_1 \cdot \mathbf{c}_3^T + 7\mathbf{a}_3 \cdot \mathbf{c}_3^T \\ &= 7(\mathbf{a}_3 \cdot \mathbf{c}_3^T) \\ &= 7 \det \mathbf{A} \\ &= 77\end{aligned}$$

□

www.RanaMaths.com

Example 12.8. Suppose that $\mathbf{A}$ is a 4×4 matrix and suppose that $\det \mathbf{A} = 11$. Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4$ denote the rows of $\mathbf{A}$. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by replacing row $\mathbf{a}_3$ by $4\mathbf{a}_1 + 5\mathbf{a}_2$, what is $\det \mathbf{B}$?

Solution. Again, this is not a Type 3 elementary row operation. The third row of $\mathbf{B}$ is $\mathbf{b}_3 = 4\mathbf{a}_1 + 5\mathbf{a}_2$. Therefore, expanding $\det \mathbf{B}$ along the third row

$$\begin{aligned}\det \mathbf{B} &= (4\mathbf{a}_1 + 5\mathbf{a}_2) \cdot \mathbf{c}_3^T \\ &= 4\mathbf{a}_1 \cdot \mathbf{c}_3^T + 5\mathbf{a}_2 \cdot \mathbf{c}_3^T \\ &= 0 + 0 \\ &= 0\end{aligned}$$

□

12.2 Determinants and Invertibility of Matrices

The following theorem characterizes invertibility of matrices with the determinant.

Theorem 12.9: A square matrix $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$.

Proof. Beginning with the matrix $\mathbf{A}$, perform elementary row operations and generate a sequence of matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p$ such that $\mathbf{A}_p$ is in row echelon form and thus triangular:

$$\mathbf{A} \sim \mathbf{A}_1 \sim \mathbf{A}_2 \sim \cdots \sim \mathbf{A}_p.$$

Thus, matrix $\mathbf{A}_i$ is obtained from $\mathbf{A}_{i-1}$ by performing one of the elementary row operations. From Theorems 12.1, 12.3, 12.4, if $\det \mathbf{A}_{i-1} \neq 0$ then $\det \mathbf{A}_i \neq 0$. In particular, $\det \mathbf{A} = 0$ if and only if $\det \mathbf{A}_p = 0$. Now, $\mathbf{A}_p$ is triangular and therefore its determinant is the product of its diagonal entries. If all the diagonal entries are non-zero then $\det \mathbf{A} = \det \mathbf{A}_p \neq 0$. In this case, $\mathbf{A}$ is invertible because there are $r = n$ leading entries in $\mathbf{A}_p$. If a diagonal entry of $\mathbf{A}_p$ is zero then $\det \mathbf{A} = \det \mathbf{A}_p = 0$. In this case, $\mathbf{A}$ is not invertible because there are $r < n$ leading entries in $\mathbf{A}_p$. Therefore, $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$. □

12.3 Properties of the Determinant

The following theorem characterizes how the determinant behaves under scalar multiplication of matrices.

Theorem 12.10: Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ and let $\mathbf{B} = \beta\mathbf{A}$, that is, $\mathbf{B}$ is obtained by multiplying every entry of $\mathbf{A}$ by β. Then $\det \mathbf{B} = \beta^n \det \mathbf{A}$.

www.RanaMaths.com

Proof. Consider the 2×2 case:

$$\begin{aligned}\det(\beta \mathbf{A}) &= \begin{vmatrix} \beta a_{11} & \beta a_{12} \\ \beta a_{12} & \beta a_{22} \end{vmatrix} \\ &= \beta a_{11} \cdot \beta a_{22} - \beta a_{12} \cdot \beta a_{21} \\ &= \beta^2 (a_{11} a_{22} - a_{12} a_{21}) \\ &= \beta^2 \det \mathbf{A}.\end{aligned}$$

Thus, the statement holds for 2×2 matrices. Consider a 3×3 matrix $\mathbf{A}$. Then

$$\begin{aligned}\det(\beta \mathbf{A}) &= \beta a_{11} |\beta \mathbf{A}_{11}| - \beta a_{12} |\beta \mathbf{A}_{12}| + \beta a_{13} |\beta \mathbf{A}_{13}| \\ &= \beta a_{11} \beta^2 |\mathbf{A}_{11}| - \beta a_{12} \beta^2 |\mathbf{A}_{12}| + \beta a_{13} \beta^2 |\mathbf{A}_{13}| \\ &= \beta^3 \left(a_{11} |\mathbf{A}_{11}| - a_{12} |\mathbf{A}_{12}| + a_{13} |\mathbf{A}_{13}| \right) \\ &= \beta^3 \det \mathbf{A}.\end{aligned}$$

The general case can be treated using mathematical induction on n. □

Example 12.11. Suppose that $\mathbf{A}$ is a 4×4 matrix and suppose that $\det \mathbf{A} = 11$. What is $\det(3\mathbf{A})$?

Solution. We have

$$\begin{aligned}\det(3\mathbf{A}) &= 3^4 \det \mathbf{A} \\ &= 81 \cdot 11 \\ &= 891\end{aligned}$$

□

The following theorem characterizes how the determinant behaves under matrix multiplication.

Theorem 12.12: Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ matrices. Then

$$\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B}).$$

Corollary 12.13: For any square matrix $\det(\mathbf{A}^k) = (\det \mathbf{A})^k$.

Corollary 12.14: If $\mathbf{A}$ is invertible then

$$\det(\mathbf{A}^{-1}) = \frac{1}{\det \mathbf{A}}.$$

Proof. From $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$ we have that $\det(\mathbf{A}\mathbf{A}^{-1}) = 1$. But also

$$\det(\mathbf{A}\mathbf{A}^{-1}) = \det(\mathbf{A})\det(\mathbf{A}^{-1}).$$

Therefore

$$\det(\mathbf{A})\det(\mathbf{A}^{-1}) = 1$$

or equivalently

$$\det \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}.$$

□

Example 12.15. Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ be $n \times n$ matrices. Suppose that $\det \mathbf{A} = 3$, $\det \mathbf{B} = 0$, and $\det \mathbf{C} = 7$.

(i) Is $\mathbf{AC}$ invertible?
(ii) Is $\mathbf{AB}$ invertible?
(iii) Is $\mathbf{ACB}$ invertible?

Solution. (i): We have $\det(\mathbf{AC}) = \det \mathbf{A} \det \mathbf{C} = 3 \cdot 7 = 21$. Thus, $\mathbf{AC}$ is invertible.
(ii): We have $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B} = 3 \cdot 0 = 0$. Thus, $\mathbf{AB}$ is not invertible.
(iii): We have $\det(\mathbf{ACB}) = \det \mathbf{A} \det \mathbf{C} \det \mathbf{B} = 3 \cdot 7 \cdot 0 = 0$. Thus, $\mathbf{ACB}$ is not invertible. □

After this lecture you should know the following:

- how the determinant behaves under elementary row operations
- that $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$
- that $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$

www.RanaMaths.com

Lecture 13

Applications of the Determinant

13.1 The Cofactor Method

Recall that for $\mathbf{A} \in \mathbb{R}^{n\times n}$ we defined

$$\det \mathbf{A} = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn}$$

where $C_{jk} = (-1)^{j+k} \det \mathbf{A}_{jk}$ is called the (j,k)-Cofactor of $\mathbf{A}$ and

$$\mathbf{a}_j = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{jn} \end{bmatrix}$$

is the jth row of $\mathbf{A}$. If $\mathbf{c}_j = \begin{bmatrix} C_{j1} & C_{j2} & \cdots & C_{jn} \end{bmatrix}$ then

$$\det \mathbf{A} = \begin{bmatrix} a_{j1} & a_{j2} & \cdots & a_{jn} \end{bmatrix} \begin{bmatrix} C_{j1} \\ C_{j2} \\ \vdots \\ C_{jn} \end{bmatrix} = \mathbf{a}_j \cdot \mathbf{c}_j^T.$$

Suppose that $\mathbf{B}$ is the matrix obtained from $\mathbf{A}$ by replacing row $\mathbf{a}_j$ with a distinct row $\mathbf{a}_k$. To compute $\det \mathbf{B}$ expand along its jth row $\mathbf{b}_j = \mathbf{a_k}$:

$$\det \mathbf{B} = \mathbf{a}_k \cdot \mathbf{c}_j^T = 0.$$

The Cofactor Method is an alternative method to find the inverse of an invertible matrix. Recall that for any matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, if we expand along the jth row then

$$\det \mathbf{A} = \mathbf{a}_j \cdot \mathbf{c}_j^T.$$

On the other hand, if $j \neq k$ then

$$\mathbf{a}_j \cdot \mathbf{c}_k^T = 0.$$

In summary,

$$\mathbf{a}_j \cdot \mathbf{c}_k^T = \begin{cases} \det \mathbf{A}, & \text{if } j = k \\ 0, & \text{if } j \neq k. \end{cases}$$

www.RanaMaths.com

Form the **Cofactor matrix**

$$\operatorname{Cof}(\mathbf{A}) = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_n \end{bmatrix}.$$

Then,

$$\begin{aligned} \mathbf{A}(\operatorname{Cof}(\mathbf{A}))^T &= \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} \begin{bmatrix} \mathbf{c}_1^T & \mathbf{c}_2^T & \cdots & \mathbf{c}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_1\mathbf{c}_1^T & \mathbf{a}_1\mathbf{c}_2^T & \cdots & \mathbf{a}_1\mathbf{c}_n^T \\ \mathbf{a}_2\mathbf{c}_1^T & \mathbf{a}_2\mathbf{c}_2^T & \cdots & \mathbf{a}_2\mathbf{c}_n^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_n\mathbf{c}_1^T & \mathbf{a}_n\mathbf{c}_2^T & \cdots & \mathbf{a}_n\mathbf{c}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \det \mathbf{A} & 0 & \cdots & 0 \\ 0 & \det \mathbf{A} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \det \mathbf{A} \end{bmatrix} \end{aligned}$$

This can be written succinctly as

$$\mathbf{A}(\operatorname{Cof}(\mathbf{A}))^T = \det(\mathbf{A})\mathbf{I}_n.$$

Now if $\det \mathbf{A} \neq 0$ then we can divide by $\det \mathbf{A}$ to obtain

$$\mathbf{A}\left(\frac{1}{\det \mathbf{A}}\right)(\operatorname{Cof}(\mathbf{A}))^T = \mathbf{I}_n.$$

This leads to the following formula for the inverse:

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}(\operatorname{Cof}(\mathbf{A}))^T$$

Although this is an explicit and elegant formula for $\mathbf{A}^{-1}$, it is computationally intensive, even for 3×3 matrices. However, for the 2×2 case it provides a useful formula to compute

www.RanaMaths.com

the matrix inverse. Indeed, if $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we have $\text{Cof}(\mathbf{A}) = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$ and therefore

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

When does an integer matrix have an integer inverse? We can answer this question using the Cofactor Method. Let us first be clear about what we mean by an integer matrix.

Definition 13.1: A matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is called an **integer matrix** if every entry of $\mathbf{A}$ is an integer.

Suppose that $\mathbf{A} \in \mathbb{R}^{n\times n}$ is an invertible integer matrix. Then $\det(\mathbf{A})$ is a non-zero integer and $(\text{Cof}(\mathbf{A}))^T$ is an integer matrix. If $\mathbf{A}^{-1}$ is also an integer matrix then $\det(\mathbf{A}^{-1})$ is also an integer. Now $\det(\mathbf{A})\det(\mathbf{A}^{-1}) = 1$ thus it must be the case that $\det(\mathbf{A}) = \pm 1$. Suppose on the other hand that $\det(\mathbf{A}) = \pm 1$. Then by the Cofactor method

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}(\text{Cof}(\mathbf{A}))^T = \pm(\text{Cof}(\mathbf{A}))^T$$

and therefore $\mathbf{A}^{-1}$ is also an integer matrix. We have proved the following.

Theorem 13.2: An invertible integer matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ has an integer inverse $\mathbf{A}^{-1}$ if and only if $\det \mathbf{A} = \pm 1$.

We can use the previous theorem to generate integer matrices with an integer inverse as follows. Begin with an upper triangular matrix $\mathbf{M}_0$ having integer entries and whose diagonal entries are either 1 or −1. By construction, $\det(\mathbf{M}_0) = \pm 1$. Perform any sequence of elementary row operations of Type 1 and Type 3. This generates a sequence of matrices $\mathbf{M}_1, \ldots, \mathbf{M}_p$ whose entries are integers. Moreover,

$$\mathbf{M}_0 \sim \mathbf{M}_1 \sim \cdots \sim \mathbf{M}_p.$$

Therefore,

$$\pm 1 = \det(\mathbf{M}) = \det(\mathbf{M}_1) = \cdots = \det(\mathbf{M}_p).$$

www.RanaMaths.com

13.2 Cramer's Rule

The Cofactor method can be used to give an explicit formula for the solution of a linear system where the coefficient matrix is invertible. The formula is known as Cramer's Rule. To derive this formula, recall that if $\mathbf{A}$ is invertible then the solution to $\mathbf{Ax} = \mathbf{b}$ is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. Using the Cofactor method, $\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}(\operatorname{Cof}(\mathbf{A}))^T$, and therefore

$$\mathbf{x} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Consider the first component x_1 of $\mathbf{x}$:

$$x_1 = \frac{1}{\det \mathbf{A}}(b_1 C_{11} + b_2 C_{21} + \cdots + b_n C_{n1}).$$

The expression $b_1 C_{11} + b_2 C_{21} + \cdots + b_n C_{n1}$ is the expansion of the determinant along the **first** column of the matrix obtained from $\mathbf{A}$ by replacing the first column with $\mathbf{b}$:

$$\det \begin{bmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{bmatrix} = b_1 C_{11} + b_2 C_{21} + \cdots + b_n C_{n1}$$

Similarly,

$$x_2 = \frac{1}{\det \mathbf{A}}(b_1 C_{12} + b_2 C_{22} + \cdots + b_n C_{n2})$$

and $(b_1 C_{12} + b_2 C_{22} + \cdots + b_n C_{n2})$ is the expansion of the determinant along the **second** column of the matrix obtained from $\mathbf{A}$ by replacing the second column with $\mathbf{b}$. In summary:

Theorem 13.3: (Cramer's Rule) Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an invertible matrix. Let $\mathbf{b} \in \mathbb{R}^n$ and let $\mathbf{A}_i$ be the matrix obtained from $\mathbf{A}$ by replacing the ith column with $\mathbf{b}$. Then the solution to $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{x} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} \det \mathbf{A}_1 \\ \det \mathbf{A}_2 \\ \vdots \\ \det \mathbf{A}_n \end{bmatrix}.$$

Although this is an explicit and elegant formula for $\mathbf{x}$, it is computationally intensive, and used mainly for theoretical purposes.

www.RanaMaths.com

13.3 Volumes

The volume of the parallelepiped determined by the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is

$$\text{Vol}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \text{abs}(\mathbf{v}_1^T(\mathbf{v}_2 \times \mathbf{v}_3)) = \text{abs}(\det \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix})$$

where abs(x) denotes the absolute value of the number x. Let $\mathbf{A}$ be an invertible matrix and let $\mathbf{w}_1 = \mathbf{A}\mathbf{v}_1, \mathbf{w}_2 = \mathbf{A}\mathbf{v}_2, \mathbf{w}_3 = \mathbf{A}\mathbf{v}_3$. How are $\text{Vol}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_2)$ and $\text{Vol}(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_2)$ related? Compute:

$$\begin{aligned}
\text{Vol}(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) &= \text{abs}(\det \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 \end{bmatrix}) \\
&= \text{abs}\left(\det \begin{bmatrix} \mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \mathbf{A}\mathbf{v}_3 \end{bmatrix}\right) \\
&= \text{abs}\left(\det(\mathbf{A} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix})\right) \\
&= \text{abs}\left(\det \mathbf{A} \cdot \det \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}\right) \\
&= \text{abs}(\det \mathbf{A}) \cdot \text{Vol}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3).
\end{aligned}$$

Therefore, the number abs$(\det \mathbf{A})$ is the factor by which volume is changed under the linear transformation with matrix $\mathbf{A}$. In summary:

Theorem 13.4: Suppose that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are vectors in $\mathbb{R}^3$ that determine a parallelepiped of non-zero volume. Let $\mathbf{A}$ be the matrix of a linear transformation and let $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ be the images of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ under $\mathbf{A}$, respectively. Then

$$\text{Vol}(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) = \text{abs}(\det \mathbf{A}) \cdot \text{Vol}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3).$$

Example 13.5. Consider the data

$$\mathbf{A} = \begin{bmatrix} 4 & 1 & -1 \\ 2 & 4 & 1 \\ 1 & 1 & 4 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -1 \\ 5 \\ 1 \end{bmatrix}.$$

and let $\mathbf{w}_1 = \mathbf{A}\mathbf{v}_1$, $\mathbf{w}_2 = \mathbf{A}\mathbf{v}_2$, and $\mathbf{w}_3 = \mathbf{A}\mathbf{v}_3$. Find the volume of the parallelepiped spanned by the vectors $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

Solution. We compute:

$$\text{Vol}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \text{abs}(\det(\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix})) = \text{abs}(-7) = 7$$

We compute:

$$\det(\mathbf{A}) = 55.$$

Therefore, the volume of the parallelepiped spanned by the vectors $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is

$$\text{Vol}(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3) = \text{abs}(55) \times 7 = 385.$$

□

www.RanaMaths.com

After this lecture you should know the following:

- what the Cofactor Method is
- what Cramer's Rule is
- the geometric interpretation of the determinant (volume)

www.RanaMaths.com

Lecture 14

Vector Spaces

14.1 Vector Spaces

When you read/hear the word **vector** you may immediately think of two points in $\mathbb{R}^2$ (or $\mathbb{R}^3$) connected by an arrow. Mathematically speaking, a **vector** is just an element of a **vector space**. This then begs the question: What is a **vector space**? Roughly speaking, a vector space is **a set of objects that can be added and multiplied by scalars**. You have already worked with several types of vector spaces. Examples of vector spaces that you have already encountered are:

1. the set $\mathbb{R}^n$,
2. the set of all $n \times n$ matrices,
3. the set of all functions from $[a, b]$ to $\mathbb{R}$, and
4. the set of all sequences.

In all of these sets, there is an operation of "addition" and "multiplication by scalars". Let's formalize then exactly what we mean by a vector space.

Definition 14.1: A **vector space** is a set V of objects, called **vectors**, on which two operations called **addition** and **scalar multiplication** have been defined satisfying the following properties. If $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are in V and if $\alpha, \beta \in \mathbb{R}$ are scalars:

(1) The sum $\mathbf{u} + \mathbf{v}$ is in V. (closure under addition)

(2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (addition is commutative)

(3) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (addition is associativity)

(4) There is a vector in V called the **zero** vector, denoted by $\mathbf{0}$, satisfying $\mathbf{v} + \mathbf{0} = \mathbf{v}$.

(5) For each $\mathbf{v}$ there is a vector $-\mathbf{v}$ in V such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

www.RanaMaths.com

(6) The scalar multiple of $\mathbf{v}$ by α, denoted $\alpha\mathbf{v}$, is in V. (closure under scalar multiplication)

(7) $\alpha(\mathbf{u}+\mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$

(8) $(\alpha+\beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$

(9) $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$

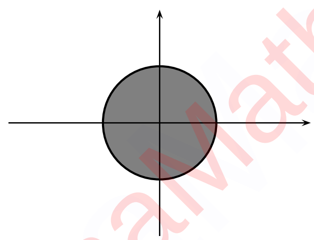
(10) $1\mathbf{v} = \mathbf{v}$

It can be shown that $0 \cdot \mathbf{v} = \mathbf{0}$ for any vector $\mathbf{v}$ in V. To better understand the definition of a vector space, we first consider a few elementary examples.

Example 14.2. Let V be the unit disc in $\mathbb{R}^2$:

$$\mathsf{V} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

Is V a vector space?

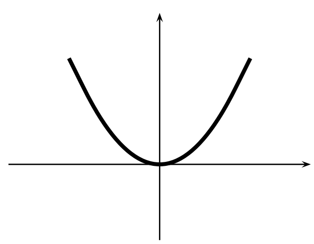


Solution. The circle is not closed under scalar multiplication. For example, take $\mathbf{u} = (1, 0) \in \mathsf{V}$ and multiply by say $\alpha = 2$. Then $\alpha\mathbf{u} = (2, 0)$ is not in V. Therefore, property (6) of the definition of a vector space fails, and consequently the unit disc is not a vector space. □

Example 14.3. Let V be the graph of the quadratic function $f(x) = x^2$:

$$\mathsf{V} = \left\{(x, y) \in \mathbb{R}^2 \mid y = x^2\right\}.$$

Is V a vector space?



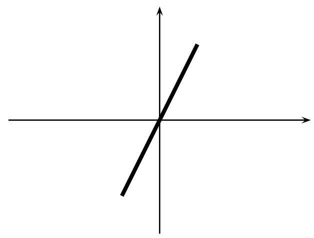
Solution. The set V is not closed under scalar multiplication. For example, $\mathbf{u} = (1, 1)$ is a point in V but $2\mathbf{u} = (2, 2)$ is not. You may also notice that V is not closed under addition either. For example, both $\mathbf{u} = (1, 1)$ and $\mathbf{v} = (2, 4)$ are in V but $\mathbf{u} + \mathbf{v} = (3, 5)$ and $(3, 5)$ is not a point on the parabola V. Therefore, the graph of $f(x) = x^2$ is not a vector space. □

www.RanaMaths.com

Example 14.4. Let V be the graph of the function $f(x) = 2x$:

$$\mathsf{V} = \{(x, y) \in \mathbb{R}^2 \mid y = 2x\}.$$

Is V a vector space?



Solution. We will show that V is a vector space. First, we verify that V is closed under addition. We first note that an arbitrary point in V can be written as $\mathbf{u} = (x, 2x)$. Let then $\mathbf{u} = (a, 2a)$ and $\mathbf{v} = (b, 2b)$ be points in V. Then

$$\mathbf{u} + \mathbf{v} = (a + b, 2a + 2b) = (a + b, 2(a + b)).$$

Therefore V is closed under addition. Verify that V is closed under scalar multiplication:

$$\alpha\mathbf{u} = \alpha(a, 2a) = (\alpha a, \alpha 2a) = (\alpha a, 2(\alpha a)).$$

Therefore V is closed under scalar multiplication. There is a zero vector $\mathbf{0} = (0, 0)$ in V:

$$\mathbf{u} + \mathbf{0} = (a, 2a) + (0, 0) = (a, 2a).$$

All the other properties of a vector space can be verified to hold; for example, addition is commutative and associative in V because addition in $\mathbb{R}^2$ is commutative/associative, etc. Therefore, the graph of the function $f(x) = 2x$ is a vector space. □

The following example is important (it will appear frequently) and is our first example of what we could say is an "abstract vector space". To emphasize, a vector space is a set that comes equipped with an operation of addition and scalar multiplication and these two operations satisfy the list of properties above.

Example 14.5. Let $\mathsf{V} = \mathbb{P}_n[t]$ be the set of all polynomials in the variable t and of degree at most n:

$$\mathbb{P}_n[t] = \left\{a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \mid a_0, a_1, \ldots, a_n \in \mathbb{R}\right\}.$$

Is V a vector space?

Solution. Let $\mathbf{u}(t) = u_0 + u_1 t + \cdots + u_n t^n$ and let $\mathbf{v}(t) = v_0 + v_1 t + \cdots + v_n t^n$ be polynomials in V. We define the addition of $\mathbf{u}$ and $\mathbf{v}$ as the new polynomial $(\mathbf{u} + \mathbf{v})$ as follows:

$$(\mathbf{u} + \mathbf{v})(t) = \mathbf{u}(t) + \mathbf{v}(t) = (u_0 + v_0) + (u_1 + v_1)t + \cdots + (u_n + v_n)t^n.$$

www.RanaMaths.com

Then $\mathbf{u}+\mathbf{v}$ is a polynomial of degree at most n and thus $(\mathbf{u}+\mathbf{v}) \in \mathbb{P}_n[t]$, and therefore this shows that $\mathbb{P}_n[t]$ is closed under addition. Now let α be a scalar, define a new polynomial $(\alpha\mathbf{u})$ as follows:

$$(\alpha\mathbf{u})(t) = (\alpha u_0) + (\alpha u_1)t + \cdots + (\alpha u_n)t^n$$

Then $(\alpha\mathbf{u})$ is a polynomial of degree at most n and thus $(\alpha\mathbf{u}) \in \mathbb{P}_n[t]$; hence, $\mathbb{P}_n[t]$ is closed under scalar multiplication. The $\mathbf{0}$ vector in $\mathbb{P}_n[t]$ is the zero polynomial $\mathbf{0}(t) = 0$. One can verify that all other properties of the definition of a vector space also hold; for example, addition is commutative and associative, etc. Thus $\mathbb{P}_n[t]$ is a vector space. □

Example 14.6. Let $\mathsf{V} = M_{m\times n}$ be the set of all $m \times n$ matrices. Under the usual operations of addition of matrices and scalar multiplication, is $M_{n\times m}$ a vector space?

Solution. Given matrices $\mathbf{A}, \mathbf{B} \in M_{m\times n}$ and a scalar α, we defined the sum $\mathbf{A}+\mathbf{B}$ by adding entry-by-entry, and $\alpha\mathbf{A}$ by multiplying each entry of $\mathbf{A}$ by α. It is clear that the space $M_{m\times n}$ is closed under these two operations. The $\mathbf{0}$ vector in $M_{m\times n}$ is the matrix of size $m \times n$ having all entries equal to zero. It can be verified that all other properties of the definition of a vector space also hold. Thus, the set $M_{m\times n}$ is a vector space. □

Example 14.7. The n-dimensional Euclidean space $\mathsf{V} = \mathbb{R}^n$ under the usual operations of addition and scalar multiplication is vector space.

Example 14.8. Let $\mathsf{V} = C[a, b]$ denote the set of functions with domain $[a, b]$ and co-domain $\mathbb{R}$ that are continuous. Is V a vector space?

14.2 Subspaces of Vector Spaces

Frequently, one encounters a vector space W that is a subset of a larger vector space V. In this case, we would say that W is a subspace of V. Below is the formal definition.

Definition 14.9: Let V be a vector space. A subset W of V is called a **subspace** of V if it satisfies the following properties:

(1) The zero vector of V is also in W.

(2) W is closed under addition, that is, if $\mathbf{u}$ and $\mathbf{v}$ are in W then $\mathbf{u}+\mathbf{v}$ is in W.

(3) W is closed under scalar multiplication, that is, if $\mathbf{u}$ is in W and α is a scalar then $\alpha\mathbf{u}$ is in W.

Example 14.10. Let W be the graph of the function $f(x) = 2x$:

$$\mathsf{W} = \{(x, y) \in \mathbb{R}^2 \mid y = 2x\}.$$

Is W a subspace of $\mathsf{V} = \mathbb{R}^2$?

www.RanaMaths.com

Solution. If $x = 0$ then $y = 2 \cdot 0 = 0$ and therefore $\mathbf{0} = (0, 0)$ is in W. Let $\mathbf{u} = (a, 2a)$ and $\mathbf{v} = (b, 2b)$ be elements of W. Then

$$\mathbf{u} + \mathbf{v} = (a, 2a) + (b, 2b) = (a + b, 2a + 2b) = (\underbrace{a + b}_{x}, 2\underbrace{(a + b)}_{x}).$$

Because the x and y components of $\mathbf{u} + \mathbf{v}$ satisfy $y = 2x$ then $\mathbf{u} + \mathbf{v}$ is inside in W. Thus, W is closed under addition. Let α be any scalar and let $\mathbf{u} = (a, 2a)$ be an element of W. Then

$$\alpha\mathbf{u} = (\alpha a, \alpha 2a) = (\underbrace{\alpha a}_{x}, 2\underbrace{(\alpha a)}_{x})$$

Because the x and y components of $\alpha\mathbf{u}$ satisfy $y = 2x$ then $\alpha\mathbf{u}$ is an element of W, and thus W is closed under scalar multiplication. All three conditions of a subspace are satisfied for W and therefore W is a subspace of V. □

Example 14.11. Let W be the first quadrant in $\mathbb{R}^2$:

$$\mathsf{W} = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0\}.$$

Is W a subspace?

Solution. The set W contains the zero vector and the sum of two vectors in W is again in W; you may want to verify this explicitly as follows: if $\mathbf{u}_1 = (x_1, y_1)$ is in W then $x_1 \geq 0$ and $y_1 \geq 0$, and similarly if $\mathbf{u}_2 = (x_2, y_2)$ is in W then $x_2 \geq 0$ and $y_2 \geq 0$. Then the sum $\mathbf{u}_1 + \mathbf{u}_2 = (x_1 + x_2, y_1 + y_2)$ has components $x_1 + y_1 \geq 0$ and $x_2 + y_2 \geq 0$ and therefore $\mathbf{u}_1 + \mathbf{u}_2$ is in W. However, W is not closed under scalar multiplication. For example if $\mathbf{u} = (1, 1)$ and $\alpha = -1$ then $\alpha\mathbf{u} = (-1, -1)$ is not in W because the components of $\alpha\mathbf{u}$ are clearly not non-negative. □

Example 14.12. Let $\mathsf{V} = M_{n\times n}$ be the vector space of all $n \times n$ matrices. We define the **trace** of a matrix $\mathbf{A} \in M_{n\times n}$ as the sum of its diagonal entries:

$$\operatorname{tr}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn}.$$

Let W be the set of all $n \times n$ matrices whose trace is zero:

$$\mathsf{W} = \{\mathbf{A} \in M_{n\times n} \mid \operatorname{tr}(\mathbf{A}) = 0\}.$$

Is W a subspace of V?

Solution. If $\mathbf{0}$ is the $n \times n$ zero matrix then clearly $\operatorname{tr}(\mathbf{0}) = 0$, and thus $\mathbf{0} \in M_{n\times n}$. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are in W. Then necessarily $\operatorname{tr}(\mathbf{A}) = 0$ and $\operatorname{tr}(\mathbf{B}) = 0$. Consider the matrix $\mathbf{C} = \mathbf{A} + \mathbf{B}$. Then

$$\begin{aligned}
\operatorname{tr}(\mathbf{C}) = \operatorname{tr}(\mathbf{A} + \mathbf{B}) &= (a_{11} + b_{11}) + (a_{22} + b_{22}) + \cdots + (a_{nn} + b_{nn}) \\
&= (a_{11} + \cdots + a_{nn}) + (b_{11} + \cdots + b_{nn}) \\
&= \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B}) \\
&= 0
\end{aligned}$$

Therefore, $\text{tr}(\mathbf{C}) = 0$ and consequently $\mathbf{C} = \mathbf{A} + \mathbf{B} \in \mathsf{W}$, in other words, W is closed under addition. Now let α be a scalar and let $\mathbf{C} = \alpha\mathbf{A}$. Then

$$\text{tr}(\mathbf{C}) = \text{tr}(\alpha\mathbf{A}) = (\alpha a_{11}) + (\alpha a_{22}) + \cdots + (\alpha a_{nn}) = \alpha\,\text{tr}(\mathbf{A}) = 0.$$

Thus, $\text{tr}(\mathbf{C}) = 0$, that is, $\mathbf{C} = \alpha\mathbf{A} \in \mathsf{W}$, and consequently W is closed under scalar multiplication. Therefore, the set W is a subspace of V. □

Example 14.13. Let $\mathsf{V} = \mathbb{P}_n[t]$ and consider the subset W of V:

$$\mathsf{W} = \{u \in \mathbb{P}_n[t] \mid u'(1) = 0\}$$

In other words, W consists of polynomials of degree n in the variable t whose derivative at $t = 1$ is zero. Is W a subspace of V?

Solution. The zero polynomial $\mathbf{0}(t) = 0$ clearly has derivative at $t = 1$ equal to zero, that is, $\mathbf{0}'(1) = 0$, and thus the zero polynomial is in W. Now suppose that $\mathbf{u}(t)$ and $\mathbf{v}(t)$ are two polynomials in W. Then, $\mathbf{u}'(1) = 0$ and also $\mathbf{v}'(1) = 0$. To verify whether or not W is closed under addition, we must determine whether the sum polynomial $(\mathbf{u} + \mathbf{v})(t)$ has a derivative at $t = 1$ equal to zero. From the rules of differentiation, we compute

$$(\mathbf{u} + \mathbf{v})'(1) = \mathbf{u}'(1) + \mathbf{v}'(1) = 0 + 0.$$

Therefore, the polynomial $(\mathbf{u} + \mathbf{v})$ is in W, and thus W is closed under addition. Now let α be any scalar and let $\mathbf{u}(t)$ be a polynomial in W. Then $\mathbf{u}'(1) = 0$. To determine whether or not the scalar multiple $\alpha\mathbf{u}(t)$ is in W we must determine if $\alpha\mathbf{u}(t)$ has a derivative of zero at $t = 1$. Using the rules of differentiation, we compute that

$$(\alpha\mathbf{u})'(1) = \alpha\mathbf{u}'(1) = \alpha \cdot 0 = 0.$$

Therefore, the polynomial $(\alpha\mathbf{u})(t)$ is in W and thus W is closed under scalar multiplication. All three properties of a subspace hold for W and therefore W is a subspace of $\mathbb{P}_n[t]$. □

Example 14.14. Let $\mathsf{V} = \mathbb{P}_n[t]$ and consider the subset W of V:

$$\mathsf{W} = \{u \in \mathbb{P}_n[t] \mid u(2) = -1\}$$

In other words, W consists of polynomials of degree n in the variable t whose value $t = 2$ is -1. Is W a subspace of V?

Solution. The zero polynomial $\mathbf{0}(t) = 0$ clearly does not equal -1 at $t = 2$. Therefore, W does not contain the zero polynomial and, because all three conditions of a subspace must be satisfied for W to be a subspace, then W is not a subspace of $\mathbb{P}_n[t]$. As an exercise, you may want to investigate whether or not W is closed under addition and scalar multiplication. □

www.RanaMaths.com

Example 14.15. A square matrix $\mathbf{A}$ is said to be **symmetric** if $\mathbf{A}^T = \mathbf{A}$. For example, here is a 3×3 symmetric matrix:

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

Verify for yourself that we do indeed have that $\mathbf{A}^T = \mathbf{A}$. Let W be the set of all symmetric $n \times n$ matrices. Is W a subspace of $\mathsf{V} = M_{n\times n}$?

Example 14.16. For any vector space V, there are two trivial subspaces in V, namely, V itself is a subspace of V and the set consisting of the zero vector $\mathsf{W} = \{\mathbf{0}\}$ is a subspace of V.

There is one particular way to generate a subspace of any given vector space V using the span of a set of vectors. Recall that we defined the span of a set of vectors in $\mathbb{R}^n$ but we can define the same notion on a general vector space V.

Definition 14.17: Let V be a vector space and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors in V. The **span** of $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$:

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} = \Big\{t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + v_p\mathbf{v}_p \mid t_1, t_2, \ldots, t_p \in \mathbb{R}\Big\}.$$

We now show that the span of a set of vectors in V is a subspace of V.

Theorem 14.18: If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are vectors in V then $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V.

Solution. Let $\mathbf{u} = t_1\mathbf{v}_1 + \cdots + t_p\mathbf{v}_p$ and $\mathbf{w} = s_1\mathbf{v}_1 + \cdots + s_p\mathbf{v}_p$ be two vectors in $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$. Then

$$\mathbf{u} + \mathbf{w} = (t_1\mathbf{v}_1 + \cdots + t_p\mathbf{v}_p) + (s_1\mathbf{v}_1 + \cdots + s_p\mathbf{v}_p) = (t_1 + s_1)\mathbf{v}_1 + \cdots + (t_p + s_p)\mathbf{v}_p.$$

Therefore $\mathbf{u} + \mathbf{w}$ is also in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_p$. Now consider $\alpha\mathbf{u}$:

$$\alpha\mathbf{u} = \alpha(t_1\mathbf{v}_1 + \cdots + t_p\mathbf{v}_p) = (\alpha t_1)\mathbf{v}_1 + \cdots + (\alpha t_p)\mathbf{v}_p.$$

Therefore, $\alpha\mathbf{u}$ is in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_p$. Lastly, since $0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p = \mathbf{0}$ then the zero vector $\mathbf{0}$ is in the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. Therefore, $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is a subspace of V. □

Given a general subspace W of V, if $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p$ are vectors in W such that

$$\text{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\} = \mathsf{W}$$

then we say that $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\}$ is a **spanning set** of W. Hence, every vector in W can be written as a linear combination of the vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p$.

After this lecture you should know the following:

www.RanaMaths.com

- what a vector space/subspace is
- be able to give some examples of vector spaces/subspaces
- that the span of a set of vectors in V is a subspace of V

Lecture 15

Linear Maps

Before we begin this Lecture, we review subspaces. Recall that W is a **subspace** of a vector space V if W is a subset of V and

1. the zero vector $\mathbf{0}$ in V is also in W,
2. for any vectors $\mathbf{u}, \mathbf{v}$ in W the sum $\mathbf{u} + \mathbf{v}$ is also in W, and
3. for any vector $\mathbf{u}$ in W and any scalar α the vector $\alpha\mathbf{u}$ is also in W.

In the previous lecture we gave several examples of subspaces. For example, we showed that a line through the origin in $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ and we gave examples of subspaces of $\mathbb{P}_n[t]$ and $M_{n\times m}$. We also showed that if $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are vectors in a vector space V then

$$\mathsf{W} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$$

is a subspace of V.

15.1 Linear Maps on Vector Spaces

In Lecture 7, we defined what it meant for a vector mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^m$ to be a **linear mapping**. We now want to introduce linear mappings on general vector spaces; you will notice that the definition is essentially the same but the key point to remember is that the underlying spaces are not $\mathbb{R}^n$ but a general vector space.

Definition 15.1: Let $\mathsf{T} : \mathsf{V} \to \mathsf{U}$ be a mapping of vector spaces. Then T is called a **linear mapping** if

- for any $\mathbf{u}, \mathbf{v}$ in V it holds that $\mathsf{T}(\mathbf{u} + \mathbf{v}) = \mathsf{T}(\mathbf{u}) + \mathsf{T}(\mathbf{v})$, and
- for any scalar α and $\mathbf{u}$ in V is holds that $\mathsf{T}(\alpha\mathbf{v}) = \alpha\mathsf{T}(\mathbf{v})$.

Example 15.2. Let $\mathsf{V} = M_{n\times n}$ be the vector space of $n \times n$ matrices and let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be the mapping

$$\mathsf{T}(\mathbf{A}) = \mathbf{A} + \mathbf{A}^T.$$

www.RanaMaths.com

Is T is a linear mapping?

Solution. Let $\mathbf{A}$ and $\mathbf{B}$ be matrices in V. Then using the properties of the transpose and regrouping we obtain:

$$\begin{aligned}\mathsf{T}(\mathbf{A}+\mathbf{B}) &= (\mathbf{A}+\mathbf{B}) + (\mathbf{A}+\mathbf{B})^T \\ &= \mathbf{A}+\mathbf{B}+\mathbf{A}^T+\mathbf{B}^T \\ &= (\mathbf{A}+\mathbf{A}^T) + (\mathbf{B}+\mathbf{B}^T) \\ &= \mathsf{T}(\mathbf{A}) + \mathsf{T}(\mathbf{B}).\end{aligned}$$

Similarly, if α is any scalar then

$$\begin{aligned}\mathsf{T}(\alpha\mathbf{A}) &= (\alpha\mathbf{A}) + (\alpha\mathbf{A})^T \\ &= \alpha\mathbf{A} + \alpha\mathbf{A}^T \\ &= \alpha(\mathbf{A}+\mathbf{A}^T) \\ &= \alpha\mathsf{T}(\mathbf{A}).\end{aligned}$$

This proves that T satisfies both conditions of Definition 15.1 and thus T is a linear mapping.

□

Example 15.3. Let $\mathsf{V} = M_{n\times n}$ be the vector space of $n \times n$ matrices, where $n \geq 2$, and let $\mathsf{T} : \mathsf{V} \to \mathbb{R}$ be the mapping

$$\mathsf{T}(\mathbf{A}) = \det(\mathbf{A})$$

Is T is a linear mapping?

Solution. **If** T is a linear mapping then according to Definition 15.1, we must have $\mathsf{T}(\mathbf{A}+\mathbf{B}) = \det(\mathbf{A}+\mathbf{B}) = \det(\mathbf{A}) + \det(\mathbf{B})$ and also $\mathsf{T}(\alpha\mathbf{A}) = \alpha\mathsf{T}(\mathbf{A})$ **for any** scalar α. Do these properties actually hold though? For example, we know from the properties of the determinant that $\det(\alpha\mathbf{A}) = \alpha^n \det(\mathbf{A})$ and therefore it does not hold that $\mathsf{T}(\alpha\mathbf{A}) = \alpha\mathsf{T}(\mathbf{A})$ unless $\alpha = 1$. Therefore, T is not a linear mapping. Also, it does not hold in general that $\det(\mathbf{A}+\mathbf{B}) = \det(\mathbf{A}) + \det(\mathbf{B})$; in fact it rarely holds. For example, if

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -1 & 1 \\ 0 & 3 \end{bmatrix}$$

then $\det(\mathbf{A}) = 2$, $\det(\mathbf{B}) = -3$ and therefore $\det(\mathbf{A}) + \det(\mathbf{B}) = -1$. On the other hand,

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 4 \end{bmatrix}$$

and thus $\det(\mathbf{A}+\mathbf{B}) = 4$. Thus, $\det(\mathbf{A}+\mathbf{B}) \neq \det(\mathbf{A}) + \det(\mathbf{B})$. □

Example 15.4. Let $\mathsf{V} = \mathbb{P}_n[t]$ be the vector space of polynomials in the variable t of degree no more than $n \geq 1$. Consider the mapping $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ define as

$$\mathsf{T}(f(t)) = 2f(t) + f'(t).$$

www.RanaMaths.com

For example, if $f(t) = 3t^6 - t^2 + 5$ then

$$\begin{aligned} \mathsf{T}(f(t)) &= 2f(t) + f'(t) \\ &= 2(3t^5 - t^2 + 5) + (18t^5 - 2t) \\ &= 6t^5 + 18t^5 - 2t^2 - 2t + 10. \end{aligned}$$

Is T is a linear mapping?

Solution. Let $f(t)$ and $g(t)$ be polynomials of degree no more than $n \geq 1$. Then

$$\begin{aligned} \mathsf{T}(f(t) + g(t)) &= 2(f(t) + g(t)) + (f(t) + g(t))' \\ &= 2f(t) + 2g(t) + f'(t) + g'(t) \\ &= (2f(t) + f'(t)) + (2g(t) + g'(t)) \\ &= \mathsf{T}(f(t)) + \mathsf{T}(g(t)). \end{aligned}$$

Therefore, $\mathsf{T}(f(t) + g(t)) = \mathsf{T}(f(t)) + \mathsf{T}(g(t))$. Now let α be any scalar. Then

$$\begin{aligned} \mathsf{T}(\alpha f(t)) &= 2(\alpha f(t)) + (\alpha f(t))' \\ &= 2\alpha f(t) + \alpha f'(t) \\ &= \alpha(2f(t) + f'(t)) \\ &= \alpha \mathsf{T}(f(t)). \end{aligned}$$

Therefore, $\mathsf{T}(\alpha f(t)) = \alpha \mathsf{T}(f(t))$. Therefore, T is a linear mapping. □

We now introduce two important subsets associated to a linear mapping.

Definition 15.5: Let $\mathsf{T} : \mathsf{V} \to \mathsf{U}$ be a linear mapping.

1. The **kernel** of T is the set of vectors $\mathbf{v}$ in the domain V that get mapped to the zero vector, that is, $\mathsf{T}(\mathbf{v}) = \mathbf{0}$. We denote the kernel of T by $\ker(\mathsf{T})$:

$$\ker(\mathsf{T}) = \{\mathbf{v} \in \mathsf{V} \mid \mathsf{T}(\mathbf{v}) = \mathbf{0}\}.$$

2. The **range** of T is the set of vectors $\mathbf{b}$ in the codomain U for which there exists at least one $\mathbf{v}$ in V such that $\mathsf{T}(\mathbf{v}) = \mathbf{b}$. We denote the range of T by $\text{Range}(\mathsf{T})$:

$$\text{Range}(\mathsf{T}) = \{\mathbf{b} \in \mathsf{U} \mid \text{there exists some } \mathbf{v} \in \mathsf{U} \text{ such that } \mathsf{T}(\mathbf{v}) = \mathbf{b}\}.$$

You may have noticed that the definition of the range of a linear mapping on an abstract vector space is the usual definition of the range of a function. Not surprisingly, the kernel and range are subspaces of the domain and codomain, respectively.

www.RanaMaths.com

Theorem 15.6: Let $\mathsf{T} : \mathsf{V} \to \mathsf{U}$ be a linear mapping. Then $\ker(\mathsf{T})$ is a subspace of V and $\text{Range}(\mathsf{T})$ is a subspace of U.

Proof. Suppose that $\mathbf{v}$ and $\mathbf{u}$ are in $\ker(\mathsf{T})$. Then $\mathsf{T}(\mathbf{v}) = \mathbf{0}$ and $\mathsf{T}(\mathbf{u}) = \mathbf{0}$. Then by linearity of T it holds that

$$\mathsf{T}(\mathbf{v} + \mathbf{u}) = \mathsf{T}(\mathbf{v}) + \mathsf{T}(\mathbf{u}) = \mathbf{0} + \mathbf{0} = \mathbf{0}.$$

Therefore, since $\mathsf{T}(\mathbf{u} + \mathbf{v}) = \mathbf{0}$ then $\mathbf{u} + \mathbf{v}$ is in $\ker(\mathsf{T})$. This shows that $\ker(\mathsf{T})$ is closed under addition. Now suppose that α is any scalar and $\mathbf{v}$ is in $\ker(\mathsf{T})$. Then $\mathsf{T}(\mathbf{v}) = \mathbf{0}$ and thus by linearity of T it holds that

$$\mathsf{T}(\alpha\mathbf{v}) = \alpha\mathsf{T}(\mathbf{v}) = \alpha\mathbf{0} = \mathbf{0}.$$

Therefore, since $\mathsf{T}(\alpha\mathbf{v}) = \mathbf{0}$ then $\alpha\mathbf{v}$ is in $\ker(\mathsf{T})$ and this proves that $\ker(\mathsf{T})$ is closed under scalar multiplication. Lastly, by linearity of T it holds that

$$\mathsf{T}(\mathbf{0}) = \mathsf{T}(\mathbf{v} - \mathbf{v}) = \mathsf{T}(\mathbf{v}) - \mathsf{T}(\mathbf{v}) = \mathbf{0}$$

that is, $\mathsf{T}(\mathbf{0}) = \mathbf{0}$. Therefore, the zero vector $\mathbf{0}$ is in $\ker(\mathsf{T})$. This proves that $\ker(\mathsf{T})$ is a subspace of V. The proof that $\text{Range}(\mathsf{T})$ is a subspace of U is left as an exercise. □

Example 15.7. Let $\mathsf{V} = M_{n\times n}$ be the vector space of $n \times n$ matrices and let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be the mapping

$$\mathsf{T}(\mathbf{A}) = \mathbf{A} + \mathbf{A}^T.$$

Describe the kernel of T.

Solution. A matrix $\mathbf{A}$ is in the kernel of T if $\mathsf{T}(\mathbf{A}) = \mathbf{A} + \mathbf{A}^T = \mathbf{0}$, that is, if $\mathbf{A}^T = -\mathbf{A}$. Hence,

$$\ker(\mathbf{A}) = \{\mathbf{A} \in M_{n\times n} \mid \mathbf{A}^T = -\mathbf{A}\}.$$

What type of matrix $\mathbf{A}$ satisfies $\mathbf{A}^T = -\mathbf{A}$? For example, consider the case that $\mathbf{A}$ is the 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and $\mathbf{A}^T = -\mathbf{A}$. Then

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

Therefore, it must hold that $a_{11} = -a_{11}$, $a_{21} = -a_{12}$ and $a_{22} = -a_{22}$. Then necessarily $a_{11} = 0$ and $a_{22} = 0$ and a_{12} can be arbitrary. For example, the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 7 \\ -7 & 0 \end{bmatrix}$$

satisfies $\mathbf{A}^T = -\mathbf{A}$. Using a similar computation as above, a 3×3 matrix satisfies $\mathbf{A}^T = -\mathbf{A}$ if $\mathbf{A}$ is of the form

$$\mathbf{A} = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

www.RanaMaths.com

where a, b, c are arbitrary constants. In general, a matrix $\mathbf{A}$ that satisfies $\mathbf{A}^T = -\mathbf{A}$ is called **skew-symmetric**. □

Example 15.8. Let V be the vector space of differentiable functions on the interval $[a, b]$. That is, f is an element of V if $f : [a, b] \to \mathbb{R}$ is differentiable. Describe the kernel of the linear mapping $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ defined as

$$\mathsf{T}(f(x)) = f(x) + f'(x).$$

Solution. A function f is in the kernel of T if $\mathsf{T}(f(x)) = 0$, that is, if $f(x) + f'(x) = 0$. Equivalently, if $f'(x) = -f(x)$. What functions f do you know of satisfy $f'(x) = -f(x)$? How about $f(x) = e^{-x}$? It is clear that $f'(x) = -e^{-x} = -f(x)$ and thus $f(x) = e^{-x}$ is in $\ker(\mathsf{T})$. How about $g(x) = 2e^{-x}$? We compute that $g'(x) = -2e^{-x} = -g(x)$ and thus g is also in $\ker(\mathsf{T})$. It turns out that the elements of $\ker(\mathsf{T})$ are of the form $f(x) = Ce^{-x}$ for a constant C. □

15.2 Null space and Column space

In the previous section, we introduced the kernel and range of a general linear mapping $\mathsf{T} : \mathsf{V} \to \mathsf{U}$. In this section, we consider the particular case of matrix mappings $\mathsf{T}_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ for some $m \times n$ matrix $\mathbf{A}$. In this case, $\mathbf{v}$ is in the kernel of $\mathsf{T}_{\mathbf{A}}$ if and only if $\mathsf{T}_{\mathbf{A}}(\mathbf{v}) = \mathbf{A}\mathbf{v} = \mathbf{0}$. In other words, $\mathbf{v} \in \ker(\mathsf{T}_{\mathbf{A}})$ if and only if $\mathbf{v}$ is a solution to the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$. Because the case when T is a matrix mapping arises so frequently, we give a name to the set of vectors $\mathbf{v}$ such that $\mathbf{A}\mathbf{v} = \mathbf{0}$.

Definition 15.9: The **null space** of a matrix $\mathbf{A} \in M_{m\times n}$, denoted by $\text{Null}(\mathbf{A})$, is the subset of $\mathbb{R}^n$ consisting of vectors $\mathbf{v}$ such that $\mathbf{A}\mathbf{v} = \mathbf{0}$. In other words, $\mathbf{v} \in \text{Null}(\mathbf{A})$ if and only if $\mathbf{A}\mathbf{v} = \mathbf{0}$. Using set notation:

$$\text{Null}(\mathbf{A}) = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{v} = \mathbf{0}\}.$$

Hence, the following holds

$$\ker(\mathsf{T}_{\mathbf{A}}) = \text{Null}(\mathbf{A}).$$

Because the kernel of a linear mapping is a subspace we obtain the following.

Theorem 15.10: If $\mathbf{A} \in M_{m\times n}$ then $\text{Null}(\mathbf{A})$ is a subspace of $\mathbb{R}^n$.

Hence, by Theorem 15.10, if $\mathbf{u}$ and $\mathbf{v}$ are two solutions to the linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ then $\alpha\mathbf{u} + \beta\mathbf{v}$ is also a solution:

$$\mathbf{A}(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha\mathbf{A}\mathbf{u} + \beta\mathbf{A}\mathbf{v} = \alpha \cdot \mathbf{0} + \beta \cdot \mathbf{0} = \mathbf{0}.$$

www.RanaMaths.com

Example 15.11. Let $\mathsf{V} = \mathbb{R}^4$ and consider the following subset of V:

$$\mathsf{W} = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid 2x_1 - 3x_2 + x_3 - 7x_4 = 0\}.$$

Is W a subspace of V?

Solution. The set W is the null space of the matrix 1×4 matrix $\mathbf{A}$ given by

$$\mathbf{A} = \begin{bmatrix} 2 & -3 & 1 & -7 \end{bmatrix}.$$

Hence, $\mathsf{W} = \text{Null}(\mathbf{A})$ and consequently W is a subspace. □

From our previous remarks, the null space of a matrix $\mathbf{A} \in M_{m\times n}$ is just the solution set of the homogeneous system $\mathbf{Ax} = \mathbf{0}$. Therefore, one way to explicitly describe the null space of $\mathbf{A}$ is to solve the system $\mathbf{Ax} = \mathbf{0}$ and write the general solution in parametric vector form. From our previous work on solving linear systems, if the $\texttt{rref}(\mathbf{A})$ has r leading 1's then the number of parameters in the solution set is $d = n - r$. Therefore, after performing back substitution, we will obtain vectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$ such that the general solution in parametric vector form can be written as

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_d\mathbf{v}_d$$

where $t_1, t_2, \ldots, t_d$ are arbitrary numbers. Therefore,

$$\text{Null}(\mathbf{A}) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d\}.$$

Hence, the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ form a spanning set for $\text{Null}(\mathbf{A})$.

Example 15.12. Find a spanning set for the null space of the matrix

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

Solution. The null space of $\mathbf{A}$ is the solution set of the homogeneous system $\mathbf{Ax} = \mathbf{0}$. Performing elementary row operations one obtains

$$\mathbf{A} \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Clearly $r = \text{rank}(\mathbf{A})$ and since $n = 5$ we will have $d = 3$ vectors in a spanning set for $\text{Null}(\mathbf{A})$. Letting $x_5 = t_1$, and $x_4 = t_2$, then from the 2nd row we obtain

$$x_3 = -2t_2 + 2t_1.$$

Letting $x_2 = t_3$, then from the 1st row we obtain

$$x_1 = 2t_3 + t_2 - 3t_1.$$

Writing the general solution in parametric vector form we obtain

$$\mathbf{x} = t_1 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + t_3 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Therefore,

$$\text{Null}(\mathbf{A}) = \text{span}\left\{ \underbrace{\begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{v}_1}, \underbrace{\begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{v}_2} \underbrace{\begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{v}_3} \right\}$$

You can verify that $\mathbf{A}\mathbf{v}_1 = \mathbf{A}\mathbf{v}_2 = \mathbf{A}\mathbf{v}_3 = \mathbf{0}$. □

Now we consider the range of a matrix mapping $\mathsf{T}_\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$. Recall that a vector $\mathbf{b}$ in the co-domain $\mathbb{R}^m$ is in the range of $\mathsf{T}_\mathbf{A}$ if there exists some vector $\mathbf{x}$ in the domain $\mathbb{R}^n$ such that $\mathsf{T}_\mathbf{A}(\mathbf{x}) = \mathbf{b}$. Since, $\mathsf{T}_\mathbf{A}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ then $\mathbf{A}\mathbf{x} = \mathbf{b}$. Now, if $\mathbf{A}$ has columns $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ then recall that

$$\mathbf{A}\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

and thus $\mathbf{A}\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{b}$. Thus, a vector $\mathbf{b}$ is in the range of $\mathbf{A}$ if it can be written as a linear combination of the columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of $\mathbf{A}$. This motivates the following definition.

Definition 15.13: Let $\mathbf{A} \in M_{m\times n}$ be a matrix. The span of the columns of $\mathbf{A}$ is called the **column space** of $\mathbf{A}$. The column space of $\mathbf{A}$ is denoted by $\text{Col}(\mathbf{A})$. Explicitly, if $\mathbf{A} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ then

$$\text{Col}(\mathbf{A}) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}.$$

In summary, we can write that

$$\text{Range}(\mathsf{T}_\mathbf{A}) = \text{Col}(\mathbf{A}).$$

and since $\text{Range}(\mathsf{T}_\mathbf{A})$ is a subspace of $\mathbb{R}^m$ then so is $\text{Col}(\mathbf{A})$.

Theorem 15.14: The column space of a $m \times n$ matrix is a subspace of $\mathbb{R}^m$.

www.RanaMaths.com

Example 15.15. Let

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & 7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}.$$

Is $\mathbf{b}$ in the column space $\text{Col}(\mathbf{A})$?

Solution. The vector $\mathbf{b}$ is in the column space of $\mathbf{A}$ if there exists $\mathbf{x} \in \mathbb{R}^4$ such that $\mathbf{Ax} = \mathbf{b}$. Hence, we must determine if $\mathbf{Ax} = \mathbf{b}$ has a solution. Performing elementary row operations on the augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ we obtain

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -2 & 1 & 3 \\ 0 & 1 & -5 & -4 & -2 \\ 0 & 0 & 0 & 17 & 1 \end{bmatrix}$$

The system is consistent and therefore $\mathbf{Ax} = \mathbf{b}$ will have a solution. Therefore, $\mathbf{b}$ is in $\text{Col}(\mathbf{A})$. □

After this lecture you should know the following:

- what the null space of a matrix is and how to compute it
- what the column space of a matrix is and how to determine if a given vector is in the column space
- what the range and kernel of a linear mapping is

Lecture 16

Linear Independence, Bases, and Dimension

16.1 Linear Independence

Roughly speaking, the concept of linear independence evolves around the idea of working with "efficient" spanning sets for a subspace. For instance, the set of directions

$$\{\text{EAST}, \text{NORTH}, \text{NORTH-EAST}\}$$

are redundant since a total displacement in the NORTH-EAST direction can be obtained by combining individual NORTH and EAST displacements. With these vague statements out of the way, we introduce the formal definition of what it means for a set of vectors to be "efficient".

Definition 16.1: Let V be a vector space and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a set of vectors in V. Then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is **linearly independent** if the only scalars $c_1, c_2, \ldots, c_p$ that satisfy the equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

are the trivial scalars $c_1 = c_2 = \cdots = c_p = 0$. If the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is not linearly independent then we say that it is **linearly dependent**.

We now describe the redundancy in a set of linear dependent vectors. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ are linearly dependent, it follows that there are scalars $c_1, c_2, \ldots, c_p$, **at least one of which is nonzero**, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}. \qquad (\star)$$

For example, suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ are linearly dependent. Then there are scalars c_1, c_2, c_3, c_4, not all of them zero, such that equation $(\star)$ holds. Suppose, for the sake of argument, that $c_3 \neq 0$. Then,

$$\mathbf{v}_3 = -\frac{c_1}{c_3}\mathbf{v}_1 - \frac{c_2}{c_3}\mathbf{v}_2 - \frac{c_4}{c_3}\mathbf{v}_4.$$

www.RanaMaths.com

Therefore, when a set of vectors is linearly dependent, it is possible to write one of the vectors as a linear combination of the others. It is in this sense that a set of linearly dependent vectors are redundant. In fact, if a set of vectors are linearly dependent we can say even more as the following theorem states.

Theorem 16.2: A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, with $\mathbf{v}_1 \neq \mathbf{0}$, is linearly dependent if and only if some $\mathbf{v}_j$ is a linear combination of the preceding vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.

Example 16.3. Show that the following set of 2×2 matrices is linearly dependent:

$$\left\{ \mathbf{A}_1 = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}, \ \mathbf{A}_2 = \begin{bmatrix} -1 & 3 \\ 1 & 0 \end{bmatrix}, \ \mathbf{A}_3 = \begin{bmatrix} 5 & 0 \\ -2 & -3 \end{bmatrix} \right\}.$$

Solution. It is clear that $\mathbf{A}_1$ and $\mathbf{A}_2$ are linearly independent, i.e., $\mathbf{A}_1$ cannot be written as a scalar multiple of $\mathbf{A}_2$, and vice-versa. Since the $(2,1)$ entry of $\mathbf{A}_1$ is zero, the only way to get the -2 in the $(2,1)$ entry of $\mathbf{A}_3$ is to multiply $\mathbf{A}_2$ by -2. Similary, since the $(2,2)$ entry of $\mathbf{A}_2$ is zero, the only way to get the -3 in the $(2,2)$ entry of $\mathbf{A}_3$ is to multiply $\mathbf{A}_1$ by 3. Hence, we suspect that $3\mathbf{A}_1 - 2\mathbf{A}_2 = \mathbf{A}_3$. Verify:

$$3\mathbf{A}_1 - 2\mathbf{A}_2 = \begin{bmatrix} 3 & 6 \\ 0 & -3 \end{bmatrix} - \begin{bmatrix} -2 & 6 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ -2 & -3 \end{bmatrix} = \mathbf{A}_3$$

Therefore, $3\mathbf{A}_1 - 2\mathbf{A}_2 - \mathbf{A}_3 = \mathbf{0}$ and thus we have found scalars c_1, c_2, c_3 not all zero such that $c_1\mathbf{A}_1 + c_2\mathbf{A}_2 + c_3\mathbf{A}_3 = \mathbf{0}$. $\square$

16.2 Bases

We now introduce the important concept of a basis. Given a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}, \mathbf{v}_p\}$ in V, we showed that $\mathsf{W} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is a subspace of V. If say $\mathbf{v}_p$ is linearly dependent on $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{p-1}$ then we can remove $\mathbf{v}_p$ and the smaller set $\{\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}\}$ still spans all of W:

$$\mathsf{W} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{p-1}, \mathbf{v}_p\} = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}\}.$$

Intuitively, $\mathbf{v}_p$ does not provide an independent "direction" in generating W. If some other vector $\mathbf{v}_j$ is linearly dependent on $\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}$ then we can remove $\mathbf{v}_j$ and the resulting smaller set of vectors still spans W. We can continue removing vectors until we obtain a minimal set of vectors that are linearly independent and still span W. The following remarks motivate the following important definition.

Definition 16.4: Let W be a subspace of a vector space V. A set of vectors $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ in W is said to be a **basis** for W if

(a) the set $\mathcal{B}$ spans all of W, that is, $\mathsf{W} = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, and

www.RanaMaths.com

(b) the set $\mathcal{B}$ is linearly independent.

A basis is therefore a **minimal** spanning set for a subspace. Indeed, if $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a basis for W and we remove say $\mathbf{v}_p$, then $\tilde{\mathcal{B}} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}\}$ cannot be a basis for W. Why? If $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a basis then it is linearly independent and therefore $\mathbf{v}_p$ **cannot** be written as a linear combination of the others. In other words, $\mathbf{v}_p \in \mathsf{W}$ is not in the span of $\tilde{\mathcal{B}} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}\}$ and therefore $\tilde{\mathcal{B}}$ is not a basis for W because a basis must be a spanning set. If, on the other hand, we start with a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ for W and we add a new vector $\mathbf{u}$ from W then $\tilde{\mathcal{B}} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{u}\}$ is not a basis for W. Why? We still have that $\operatorname{span}\tilde{\mathcal{B}} = \mathsf{W}$ but now $\tilde{\mathcal{B}}$ is not linearly independent. Indeed, because $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a basis for W, the vector $\mathbf{u}$ can be written as a linear combination of $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, and thus $\tilde{\mathcal{B}}$ is not linearly independent.

Example 16.5. Show that the standard unit vectors form a basis for $\mathsf{V} = \mathbb{R}^3$:

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \quad \mathbf{e}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$$

Solution. Any vector $\mathbf{x} \in \mathbb{R}^3$ can be written as a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$:

$$\mathbf{x} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = x_1\begin{bmatrix}1\\0\\0\end{bmatrix} + x_2\begin{bmatrix}0\\1\\0\end{bmatrix} + x_3\begin{bmatrix}0\\0\\1\end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$$

Therefore, $\operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \mathbb{R}^3$. The set $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is linearly independent. Indeed, if there are scalars c_1, c_2, c_3 such that

$$c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 = \mathbf{0}$$

then clearly they must all be zero, $c_1 = c_2 = c_3 = 0$. Therefore, by definition, $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a basis for $\mathbb{R}^3$. This basis is called the **standard basis** for $\mathbb{R}^3$. Analogous arguments hold for $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ in $\mathbb{R}^n$. □

Example 16.6. Is $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ a basis for $\mathbb{R}^3$?

$$\mathbf{v}_1 = \begin{bmatrix}2\\0\\-4\end{bmatrix}, \; \mathbf{v}_2 = \begin{bmatrix}-4\\-2\\8\end{bmatrix}, \; \mathbf{v}_3 = \begin{bmatrix}4\\-6\\-6\end{bmatrix}$$

Solution. Form the matrix $\mathbf{A} = [\mathbf{v}_1 \; \mathbf{v}_2 \; \mathbf{v}_3]$ and row reduce:

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

Therefore, the only solution to $\mathbf{Ax} = \mathbf{0}$ is the trivial solution. Therefore, $\mathcal{B}$ is linearly independent. Moreover, for any $\mathbf{b} \in \mathbb{R}^3$, the augmented matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ is consistent. Therefore, the columns of $\mathbf{A}$ span all of $\mathbb{R}^3$:

$$\text{Col}(\mathbf{A}) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \mathbb{R}^3.$$

Therefore, $\mathcal{B}$ is a basis for $\mathbb{R}^3$. □

Example 16.7. In $\mathsf{V} = \mathbb{R}^4$, consider the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ -1 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 4 \\ 2 \\ -3 \end{bmatrix}.$$

Let $\mathsf{W} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Is $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ a basis for W?

Solution. By definition, $\mathcal{B}$ is a spanning set for W, so we need only determine if $\mathcal{B}$ is linearly independent. Form the matrix, $\mathbf{A} = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]$ and row reduce to obtain

$$\mathbf{A} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, $\text{rank}(\mathbf{A}) = 2$ and thus $\mathcal{B}$ is linearly dependent. Notice $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{v}_3$. Therefore, $\mathcal{B}$ is not a basis of W. □

Example 16.8. Find a basis for the vector space of 2×2 matrices.

Example 16.9. Recall that a $n \times n$ is skew-symmetric $\mathbf{A}$ if $\mathbf{A}^T = -\mathbf{A}$. We proved that the set of $n \times n$ matrices is a subspace. Find a basis for the set of 3×3 skew-symmetric matrices.

16.3 Dimension of a Vector Space

The following theorem will lead to the definition of the dimension of a vector space.

Theorem 16.10: Let V be a vector space. Then all bases of V have the same number of vectors.

Proof: We will prove the theorem for the case that $\mathsf{V} = \mathbb{R}^n$. We already know that the standard unit vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis of $\mathbb{R}^n$. Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be nonzero vectors in $\mathbb{R}^n$ and suppose first that $p > n$. In Lecture 6, Theorem 6.7, we proved that any set of vectors in $\mathbb{R}^n$ containing more than n vectors is automatically linearly dependent. The reason is that the RREF of $\mathbf{A} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix}$ will contain at most $r = n$ leading ones,

www.RanaMaths.com

and therefore $d = p - n > 0$. Therefore, the solution set of $\mathbf{Ax} = \mathbf{0}$ contains non-trivial solutions. On the other hand, suppose instead that $p < n$. In Lecture 4, Theorem 4.11, we proved that a set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ spans $\mathbb{R}^n$ if and only if the RREF of $\mathbf{A}$ has exactly $r = n$ leading ones. The largest possible value of r is $r = p < n$. Therefore, if $p < n$ then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ cannot be a basis for $\mathbb{R}^n$. Thus, in either case ($p > n$ or $p < n$), the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ cannot be a basis for $\mathbb{R}^n$. Hence, any basis in $\mathbb{R}^n$ must contain n vectors. □

The previous theorem does not say that every set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of nonzero vectors in $\mathbb{R}^n$ containing n vectors is automatically a basis for $\mathbb{R}^n$. For example,

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$$

do not form a basis for $\mathbb{R}^3$ because

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

is not in the span of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. All that we can say is that a set of vectors in $\mathbb{R}^n$ containing fewer or more than n vectors is automatically **not** a basis for $\mathbb{R}^n$. From Theorem 16.10, any basis in $\mathbb{R}^n$ must have exactly n vectors. In fact, on a general abstract vector space V, if $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V then any other basis for V must have exactly n vectors also. Because of this result, we can make the following definition.

Definition 16.11: Let V be a vector space. The **dimension** of V, denoted $\dim \mathsf{V}$, is the number of vectors in any basis of V. The dimension of the trivial vector space $\mathsf{V} = \{\mathbf{0}\}$ is defined to be zero.

There is one subtle issue we are sweeping under the rug: Does every vector space have a basis? The answer is yes but we will not prove this result here.

Moving on, suppose that we have a set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in $\mathbb{R}^n$ containing exactly n vectors. For $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ to be a basis of $\mathbb{R}^n$, the set $\mathcal{B}$ must be linearly independent and $\operatorname{span} \mathcal{B} = \mathbb{R}^n$. In fact, it can be shown that if $\mathcal{B}$ is linearly independent then the spanning condition $\operatorname{span} \mathcal{B} = \mathbb{R}^n$ is automatically satisfied, and vice-versa. For example, say the vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in $\mathbb{R}^n$ are linearly independent, and put $\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$. Then $\mathbf{A}^{-1}$ exists and therefore $\mathbf{Ax} = \mathbf{b}$ is always solvable. Hence, $\operatorname{Col}(\mathbf{A}) = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \mathbb{R}^n$. In summary, we have the following theorem.

Theorem 16.12: Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be vectors in $\mathbb{R}^n$. If $\mathcal{B}$ is linearly independent then $\mathcal{B}$ is a basis for $\mathbb{R}^n$. Or if $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = \mathbb{R}^n$ then $\mathcal{B}$ is a basis for $\mathbb{R}^n$.

www.RanaMaths.com

Example 16.13. Do the columns of the matrix $\mathbf{A}$ form a basis for $\mathbb{R}^4$?

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 3 & -2 \\ 4 & 7 & 8 & -6 \\ 0 & 0 & 1 & 0 \\ -4 & -6 & -6 & 3 \end{bmatrix}$$

Solution. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ denote the columns of $\mathbf{A}$. Since we have $n = 4$ vectors in $\mathbb{R}^n$, we need only check that they are linearly independent. Compute

$$\det \mathbf{A} = -2 \neq 0$$

Hence, $\operatorname{rank}(\mathbf{A}) = 4$ and thus the columns of $\mathbf{A}$ are linearly independent. Therefore, the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ form a basis for $\mathbb{R}^4$. □

A subspace W of a vector space V is a vector space in its own right, and therefore also has dimension. By definition, if $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly independent set in W and $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathsf{W}$, then $\mathcal{B}$ is a basis for W and in this case the dimension of W is k. Since an n-dimensional vector space V requires exactly n vectors in any basis, then if W is a strict subspace of V then

$$\dim \mathsf{W} < \dim \mathsf{V}.$$

As an example, in $\mathsf{V} = \mathbb{R}^3$ subspaces can be classified by dimension:

1. The zero dimensional subspace in $\mathbb{R}^3$ is $\mathsf{W} = \{\mathbf{0}\}$.
2. The one dimensional subspaces in $\mathbb{R}^3$ are lines through the origin. These are spanned by a single non-zero vector.
3. The two dimensional subspaces in $\mathbb{R}^3$ are planes through the origin. These are spanned by two linearly independent vectors.
4. The only three dimensional subspace in $\mathbb{R}^3$ is $\mathbb{R}^3$ itself. Any set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in $\mathbb{R}^3$ that is linearly independent is a basis for $\mathbb{R}^3$.

Example 16.14. Find a basis for $\operatorname{Null}(\mathbf{A})$ and the $\dim \operatorname{Null}(\mathbf{A})$ if

$$\mathbf{A} = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix}.$$

Solution. By definition, the $\operatorname{Null}(\mathbf{A})$ is the solution set of the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$. Row reducing we obtain

$$\mathbf{A} \sim \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & 5/2 & 3/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

www.RanaMaths.com

The general solution to $\mathbf{Ax} = \mathbf{0}$ in parametric form is

$$\mathbf{x} = t \begin{bmatrix} -5 \\ -3/2 \\ 0 \\ 1 \end{bmatrix} + s \begin{bmatrix} -6 \\ -5/2 \\ 1 \\ 0 \end{bmatrix} = t\mathbf{v}_1 + s\mathbf{v}_2$$

By construction, the vectors

$$\mathbf{v}_1 = \begin{bmatrix} -5 \\ -3/2 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -6 \\ -5/2 \\ 1 \\ 0 \end{bmatrix}$$

span the null space $(\mathbf{A})$ and they are linearly independent. Therefore, $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\text{Null}(\mathbf{A})$ and therefore $\dim \text{Null}(\mathbf{A}) = 2$. In general, the dimension of the $\text{Null}(\mathbf{A})$ is the number of free parameters in the solution set of the system $\mathbf{Ax} = \mathbf{0}$, that is,

$$\dim \text{Null}(\mathbf{A}) = d = n - \text{rank}(\mathbf{A})$$

□

Example 16.15. Find a basis for $\text{Col}(\mathbf{A})$ and the $\dim \text{Col}(\mathbf{A})$ if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & -4 & 8 \\ 1 & 2 & 0 & 2 & 8 \\ 2 & 4 & -3 & 10 & 9 \\ 3 & 6 & 0 & 6 & 9 \end{bmatrix}.$$

Solution. By definition, the column space of $\mathbf{A}$ is the span of the columns of $\mathbf{A}$, which we denote by $\mathbf{A} = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3\ \mathbf{v}_4\ \mathbf{v}_5]$. Thus, to find a basis for $\text{Col}(\mathbf{A})$, by trial and error we could determine the largest subset of the columns of $\mathbf{A}$ that are linearly independent. For example, first we determine if $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent. If yes, then add $\mathbf{v}_3$ and determine if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent. If $\{\mathbf{v}_1, \mathbf{v}_2\}$ is not linearly independent then discard $\mathbf{v}_2$ and determine if $\{\mathbf{v}_1, \mathbf{v}_3\}$ is linearly independent. We continue this process until we have determined the largest subset of the columns of $\mathbf{A}$ that is linearly independent, and this will yield a basis for $\text{Col}(\mathbf{A})$. Instead, we can use the fact that matrices that are row equivalent induce the same solution set for the associated homogeneous system. Hence, let $\mathbf{B}$ be the RREF of $\mathbf{A}$:

$$\mathbf{B} = \text{rref}(\mathbf{A}) = \begin{bmatrix} 1 & 2 & 0 & 2 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

www.RanaMaths.com

By inspection, the columns $\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5$ of $\mathbf{B}$ are linearly independent. It is easy to see that $\mathbf{b}_2 = 2\mathbf{b}_1$ and $\mathbf{b}_4 = 2\mathbf{b}_1 - 2\mathbf{b}_3$. These same linear relations hold for the columns of $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & -4 & 8 \\ 1 & 2 & 0 & 2 & 8 \\ 2 & 4 & -3 & 10 & 9 \\ 3 & 6 & 0 & 6 & 9 \end{bmatrix}$$

By inspection, $\mathbf{v}_2 = 2\mathbf{v}_1$ and $\mathbf{v}_4 = 2\mathbf{v}_1 - 2\mathbf{v}_3$. Thus, because $\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5$ are linearly independent columns of $\mathbf{B} = \texttt{rref}(\mathbf{A})$, then $\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_5$ are linearly independent columns of $\mathbf{A}$. Therefore, we have

$$\text{Col}(\mathbf{A}) = \text{span}\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_5\} = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 8 \\ 8 \\ 9 \\ 9 \end{bmatrix} \right\}$$

and consequently $\dim \text{Col}(\mathbf{A}) = 3$. **This procedure works in general:** To find a basis for the $\text{Col}(\mathbf{A})$, row reduce $\mathbf{A} \sim \mathbf{B}$ until you can determine which columns of $\mathbf{B}$ are linearly independent. The columns of $\mathbf{A}$ in the same position as the linearly independent columns of $\mathbf{B}$ form a basis for the $\text{Col}(\mathbf{A})$.

WARNING: Do not take the linearly independent columns of $\mathbf{B}$ as a basis for $\text{Col}(\mathbf{A})$. Always go back to the original matrix $\mathbf{A}$ to select the columns. □

After this lecture you should know the following:

- what it means for a set to be linearly independent/dependents
- what a basis is (a spanning set that is linearly independent)
- what is the meaning of the dimension of a vector space
- how to determine if a given set in $\mathbb{R}^n$ is linearly independent
- how to find a basis for the null space and column space of a matrix $\mathbf{A}$

www.RanaMaths.com

Lecture 17

The Rank Theorem

17.1 The Rank of a Matrix

We now give the definition to the rank of a matrix.

Definition 17.1: The **rank** of a matrix $\mathbf{A}$ is the dimension of its column space. We will use $\text{rank}(\mathbf{A})$ to denote the rank of $\mathbf{A}$.

Recall that $\text{Col}(\mathbf{A}) = \text{Range}(\mathsf{T}_\mathbf{A})$, and thus the rank of $\mathbf{A}$ is the dimension of the range of the linear mapping $\mathsf{T}_\mathbf{A}$. The range of a mapping is sometimes called the **image**.

We now define the nullity of a matrix.

Definition 17.2: The **nullity** of a matrix $\mathbf{A}$ is the dimension of its nullspace $\text{Null}(\mathbf{A})$. We will use $\text{nullity}(\mathbf{A})$ to denote the nullity of $\mathbf{A}$.

Recall that $(\mathbf{A}) = \ker(\mathsf{T}_\mathbf{A})$, and thus the nullity of $\mathbf{A}$ is the dimension of the kernel of the linear mapping $\mathsf{T}_\mathbf{A}$.

The rank and nullity of a matrix are connected via the following fundamental theorem known as the Rank Theorem.

Theorem 17.3: (Rank Theorem) Let $\mathbf{A}$ be a $m \times n$ matrix. The rank of $\mathbf{A}$ is the number of leading 1's in its RREF. Moreover, the following equation holds:

$$n = \text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}).$$

Proof. A basis for the column space is obtained by computing $\texttt{rref}(\mathbf{A})$ and identifying the columns that contain a leading 1. Each column of $\mathbf{A}$ corresponding to a column of $\texttt{rref}(\mathbf{A})$ with a leading 1 is a basis vector for the column space of $\mathbf{A}$. Therefore, if r is the number of leading 1's then $r = \text{rank}(\mathbf{A})$. Now let $d = n - r$. The number of free parameters in the

www.RanaMaths.com

solution set of $\mathbf{Ax} = \mathbf{0}$ is d and therefore a basis for $\text{Null}(\mathbf{A})$ will contain d vectors, that is, $\text{nullity}(\mathbf{A}) = d$. Therefore,

$$\text{nullity}(\mathbf{A}) = n - \text{rank}(\mathbf{A}).$$

□

Example 17.4. Find the rank and nullity of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 2 & 3 & -6 \\ 0 & -1 & -3 & 1 & 1 \\ -2 & 4 & -3 & -6 & 11 \end{bmatrix}.$$

Solution. Row reduce far enough to identify where the leading entries are:

$$\mathbf{A} \xrightarrow{2R_1+R_2} \begin{bmatrix} 1 & -2 & 2 & 3 & -6 \\ 0 & -1 & -3 & 1 & 1 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

There are $r = 3$ leading entries and therefore $\text{rank}(\mathbf{A}) = 3$. The nullity is therefore $\text{nullity}(\mathbf{A}) = 5 - \text{rank}(\mathbf{A}) = 2$. □

Example 17.5. Find the rank and nullity of the matrix

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

Solution. Row reduce far enough to identify where the leading entries are:

$$\mathbf{A} \xrightarrow{R_1+R_2, R_1+R_3} \begin{bmatrix} 1 & -3 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

There are $r = 3$ leading entries and therefore $\text{rank}(\mathbf{A}) = 3$. The nullity is therefore $\text{nullity}(\mathbf{A}) = 3 - \text{rank}(\mathbf{A}) = 0$. Another way to see that $\text{nullity}(\mathbf{A}) = 0$ is as follows. From the above computation, $\mathbf{A}$ is invertible. Therefore, there is only one vector in $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$. The subspace $\{\mathbf{0}\}$ has dimension zero. □

Using the rank and nullity of a matrix, we now provide further characterizations of invertible matrices.

Theorem 17.6: Let $\mathbf{A}$ be a $n \times n$ matrix. The following statements are equivalent:

(i) The columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.

(ii) $\text{Col}(\mathbf{A}) = \mathbb{R}^n$

(iii) $\text{rank}(\mathbf{A}) = n$

(iv) $\text{Null}(\mathbf{A}) = \{\mathbf{0}\}$

(v) nullity($\mathbf{A}$) $= 0$

(vi) $\mathbf{A}$ is an invertible matrix.

After this lecture you should know the following:

- what the rank of a matrix is and how to compute it
- what the nullity of a matrix is and how to compute it
- the Rank Theorem

www.RanaMaths.com

www.RanaMaths.com

Lecture 18

Coordinate Systems

18.1 Coordinates

Recall that a basis of a vector space V is a set of vectors $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in V such that

1. the set $\mathcal{B}$ spans all of V, that is, $\mathsf{V} = \operatorname{span}(\mathcal{B})$, and
2. the set $\mathcal{B}$ is linearly independent.

Hence, if $\mathcal{B}$ is a basis for V, each vector $\mathbf{x}^* \in \mathsf{V}$ can be written as a linear combination of $\mathcal{B}$:

$$\mathbf{x}^* = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Moreover, from the definition of linear independence given in Definition 6.1, any vector $\mathbf{x} \in \operatorname{span}(\mathcal{B})$ can be written in **only one** way as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. In other words, for the $\mathbf{x}^*$ above, there does not exist other scalars $t_1, \ldots, t_n$ such that also

$$\mathbf{x}^* = t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n.$$

To see this, suppose that we can write $\mathbf{x}^*$ in two different ways using $\mathcal{B}$:

$$\begin{aligned}\mathbf{x}^* &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n\\ \mathbf{x}^* &= t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_n\mathbf{v}_n.\end{aligned}$$

Then

$$\mathbf{0} = \mathbf{x}^* - \mathbf{x}^* = (c_1 - t_1)\mathbf{v}_1 + (c_2 - t_2)\mathbf{v}_2 + \cdots + (c_n - t_n)\mathbf{v}_n.$$

Since $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent, the only linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ that gives the zero vector $\mathbf{0}$ is the trivial linear combination. Therefore, it must be the case that $c_i - t_i = 0$, or equivalently that $c_i = t_i$ for all $i = 1, 2 \ldots, n$. Thus, there is only one way to write $\mathbf{x}^*$ in terms of $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Hence, relative to the basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, the scalars $c_1, c_2, \ldots, c_n$ uniquely determine the vector $\mathbf{x}$, and vice-versa.

Our preceding discussion on the unique representation property of vectors in a given basis leads to the following definition.

www.RanaMaths.com

Definition 18.1: Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for V and let $\mathbf{x} \in \mathsf{V}$. The **coordinates of x relative to the basis** $\mathcal{B}$ are the unique scalars $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

In vector notation, the $\mathcal{B}$-coordinates of $\mathbf{x}$ will be denoted by

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

and we will call $[\mathbf{x}]_{\mathcal{B}}$ the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$.

The notation $[\mathbf{x}]_{\mathcal{B}}$ indicates that these are coordinates of $\mathbf{x}$ with respect to the basis $\mathcal{B}$. If it is clear what basis we are working with, we will omit the subscript $\mathcal{B}$ and simply write $[\mathbf{x}]$ for the coordinates of $\mathbf{x}$ relative to $\mathcal{B}$.

Example 18.2. One can verify that

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^2$. Find the coordinates of $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ relative to $\mathcal{B}$.

Solution. Let $\mathbf{v}_1 = (1, 1)$ and let $\mathbf{v}_2 = (-1, 1)$. By definition, the coordinates of $\mathbf{v}$ with respect to $\mathcal{B}$ are the scalars c_1, c_2 such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

If we put $\mathbf{P} = [\mathbf{v}_1\ \mathbf{v}_2]$, and let $[\mathbf{v}]_{\mathcal{B}} = (c_1, c_2)$, then we need to solve the linear system

$$\mathbf{v} = \mathbf{P}[\mathbf{v}]_{\mathcal{B}}$$

Solving the linear system, one finds that the solution is $[\mathbf{v}]_{\mathcal{B}} = (2, -1)$, and therefore this is the $\mathcal{B}$-coordinate vector of $\mathbf{v}$, or the coordinates of $\mathbf{v}$, relative to $\mathcal{B}$. □

It is clear how the procedure of the previous example can be generalized. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$ and let $\mathbf{v}$ be any vector in $\mathbb{R}^n$. Put $\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$. Then the $\mathcal{B}$-coordinates of $\mathbf{v}$ is the unique column vector $[\mathbf{v}]_{\mathcal{B}}$ solving the linear system

$$\mathbf{P}\mathbf{x} = \mathbf{v}$$

www.RanaMaths.com

that is, $\mathbf{x} = [\mathbf{v}]_{\mathcal{B}}$ is the unique solution to $\mathbf{Px} = \mathbf{v}$. Because $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent, the solution to $\mathbf{Px} = \mathbf{v}$ is

$$[\mathbf{v}]_{\mathcal{B}} = \mathbf{P}^{-1}\mathbf{v}.$$

We remark that if an inconsistent row arises when you row reduce the augmented matrix $\begin{bmatrix}\mathbf{P} & \mathbf{v}\end{bmatrix}$ then you have made an error in your row reduction algorithm. In summary, to find coordinates with respect to a basis $\mathcal{B}$ in $\mathbb{R}^n$, we need to solve a square linear system.

Example 18.3. Let

$$\mathbf{v}_1 = \begin{bmatrix}3\\6\\2\end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix}-1\\0\\1\end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix}3\\12\\7\end{bmatrix}$$

and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$. One can show that $\mathcal{B}$ is linearly independent and therefore a basis for $\mathsf{W} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Determine if $\mathbf{x}$ is in W, and if so, find the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$.

Solution. By definition, $\mathbf{x}$ is in $\mathsf{W} = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ if we can write $\mathbf{x}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$:

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$$

Form the associated augmented matrix and row reduce:

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

The system is consistent with solution $c_1 = 2$ and $c_2 = 3$. Therefore, $\mathbf{x}$ is in W, and the $\mathcal{B}$-coordinates of $\mathbf{x}$ are

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix}2\\3\end{bmatrix}$$

□

Example 18.4. What are the coordinates of

$$\mathbf{v} = \begin{bmatrix}3\\11\\-7\end{bmatrix}$$

in the standard basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$?

Solution. Clearly,

$$\mathbf{v} = \begin{bmatrix}3\\11\\-7\end{bmatrix} = 3\begin{bmatrix}1\\0\\0\end{bmatrix} + 11\begin{bmatrix}0\\1\\0\end{bmatrix} - 7\begin{bmatrix}0\\0\\1\end{bmatrix}$$

www.RanaMaths.com

Therefore, the coordinate vector of $\mathbf{v}$ relative to $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is

$$[\mathbf{v}]_{\mathcal{E}} = \begin{bmatrix} 3 \\ 11 \\ -7 \end{bmatrix}$$

□

Example 18.5. Let $\mathbb{P}_3[t]$ be the vector space of polynomials of degree at most 3.

(i) Show that $\mathcal{B} = \{1, t, t^2, t^3\}$ is a basis for $\mathbb{P}_3[t]$.

(ii) Find the coordinates of $v(t) = 3 - t^2 - 7t^3$ relative to $\mathcal{B}$.

Solution. The set $\mathcal{B} = \{1, t, t^2, t^3\}$ is a spanning set for $\mathbb{P}_3[t]$. Indeed, any polynomial $u(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$ is clearly a linear combination of $1, t, t^2, t^3$. Is $\mathcal{B}$ linearly independent? Suppose that there exists scalars c_0, c_1, c_2, c_3 such that

$$c_0 + c_1 t + c_2 t^2 + c_3 t^3 = 0.$$

Since the above equality must hold for all values of t, we conclude that $c_0 = c_1 = c_2 = c_3 = 0$. Therefore, $\mathcal{B}$ is linearly independent, and consequently a basis for $\mathbb{P}_3[t]$. In the basis $\mathcal{B}$, the coordinates of $v(t) = 3 - t^2 - 7t^3$ are

$$[v(t)]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 0 \\ -1 \\ -7 \end{bmatrix}$$

The basis $\mathcal{B} = \{1, t, t^2, t^3\}$ is called the **standard basis** in $\mathbb{P}_3[t]$. □

Example 18.6. Show that

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is a basis for $M_{2\times 2}$. Find the coordinates of $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ -4 & -1 \end{bmatrix}$ relative to $\mathcal{B}$.

Solution. Any matrix $\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$ can be written as a linear combination of the matrices in $\mathcal{B}$:

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = m_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + m_{12} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + m_{21} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + m_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

If

$$c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

www.RanaMaths.com

then clearly $c_1 = c_2 = c_3 = c_4 = 0$. Therefore, $\mathcal{B}$ is linearly independent, and consequently a basis for $M_{2\times 2}$. The coordinates of $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ -4 & -1 \end{bmatrix}$ in the basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

are

$$[\mathbf{A}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 0 \\ -4 \\ -1 \end{bmatrix}$$

The basis $\mathcal{B}$ above is the **standard basis** of $M_{2\times 2}$. □

18.2 Coordinate Mappings

Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of $\mathbb{R}^n$ and let $\mathbf{P} = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] \in M_{n\times n}$. If $\mathbf{x} \in \mathbb{R}^n$ and $[\mathbf{x}]_{\mathcal{B}}$ are the $\mathcal{B}$-coordinates of $\mathbf{x}$ relative to $\mathcal{B}$ then

$$\mathbf{x} = \mathbf{P}[\mathbf{x}]_{\mathcal{B}}. \qquad (\star)$$

Hence, thinking of $\mathbf{P} : \mathbb{R}^n \to \mathbb{R}^n$ as a linear mapping, $\mathbf{P}$ maps $\mathcal{B}$-coordinate vectors to coordinate vectors relative to the standard basis of $\mathbb{R}^n$. For this reason, we call $\mathbf{P}$ the **change-of-coordinates matrix** from the basis $\mathcal{B}$ to the standard basis in $\mathbb{R}^n$. If we need to emphasize that $\mathbf{P}$ is constructed from the basis $\mathcal{B}$ we will write $\mathbf{P}_{\mathcal{B}}$ instead of just $\mathbf{P}$. Multiplying equation ($\star$) by $\mathbf{P}^{-1}$ we obtain

$$\mathbf{P}^{-1}\mathbf{x} = [\mathbf{x}]_{\mathcal{B}}.$$

Therefore, $\mathbf{P}^{-1}$ maps coordinate vectors in the standard basis to coordinates relative to $\mathcal{B}$.

Example 18.7. The columns of the matrix $\mathbf{P}$ form a basis $\mathcal{B}$ for $\mathbb{R}^3$:

$$\mathbf{P} = \begin{bmatrix} 1 & 3 & 3 \\ -1 & -4 & -2 \\ 0 & 0 & -1 \end{bmatrix}.$$

(a) What vector $\mathbf{x} \in \mathbb{R}^3$ has $\mathcal{B}$-coordinates $[\mathbf{x}]_{\mathcal{B}} = (1, 0, -1)$.

(b) Find the $\mathcal{B}$-coordinates of $\mathbf{v} = (2, -1, 0)$.

Solution. The matrix $\mathbf{P}$ maps $\mathcal{B}$-coordinates to standard coordinates in $\mathbb{R}^3$. Therefore,

$$\mathbf{x} = \mathbf{P}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

www.RanaMaths.com

On the other hand, the inverse matrix $\mathbf{P}^{-1}$ maps standard coordinates in $\mathbb{R}^3$ to $\mathcal{B}$-coordinates. One can verify that

$$\mathbf{P}^{-1} = \begin{bmatrix} 4 & 3 & 6 \\ -1 & -1 & -1 \\ 0 & 0 & -1 \end{bmatrix}$$

Therefore, the $\mathcal{B}$ coordinates of $\mathbf{v}$ are

$$[\mathbf{v}]_{\mathcal{B}} = \mathbf{P}^{-1}\mathbf{v} = \begin{bmatrix} 4 & 3 & 6 \\ -1 & -1 & -1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 0 \end{bmatrix}$$

□

When V is an abstract vector space, e.g. $\mathbb{P}_n[t]$ or $M_{n\times n}$, the notion of a coordinate mapping is similar as the case when $\mathsf{V} = \mathbb{R}^n$. If V is an n-dimensional vector space and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, we define the coordinate mapping $\mathcal{P} : \mathsf{V} \to \mathbb{R}^n$ relative to $\mathcal{B}$ as the mapping

$$\mathcal{P}(\mathbf{v}) = [\mathbf{v}]_{\mathcal{B}}.$$

Example 18.8. Let $\mathsf{V} = M_{2\times 2}$ and let $\mathcal{B} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\}$ be the standard basis for $M_{2\times 2}$. What is $\mathcal{P} : M_{2\times 2} \to \mathbb{R}^4$?

Solution. Recall,

$$\mathcal{B} = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Then for any $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ we have

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

□

18.3 Matrix Representation of a Linear Map

Let V and W be vector spaces and let $\mathsf{T} : \mathsf{V} \to \mathsf{W}$ be a linear mapping. Then by definition of a linear mapping, $\mathsf{T}(\mathbf{v} + \mathbf{u}) = \mathsf{T}(\mathbf{v}) + \mathsf{T}(\mathbf{u})$ and $\mathsf{T}(\alpha\mathbf{v}) = \alpha\mathsf{T}(\mathbf{v})$ for every $\mathbf{v}, \mathbf{u} \in \mathsf{V}$ and $\alpha \in \mathbb{R}$. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of V and let $\gamma = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be a basis of W. Then for any $\mathbf{v} \in \mathsf{V}$ there exists scalars $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

www.RanaMaths.com

and thus $[\mathbf{v}]_{\mathcal{B}} = (c_1, c_2, \ldots, c_n)$ are the coordinates of $\mathbf{v}$ in the basis $\mathcal{B}$ By linearity of the mapping T we have

$$\mathsf{T}(\mathbf{v}) = \mathsf{T}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n)$$

$$= c_1\mathsf{T}(\mathbf{v}_1) + c_2\mathsf{T}(\mathbf{v}_2) + \cdots + c_n\mathsf{T}(\mathbf{v}_n)$$

Now each vector $\mathsf{T}(\mathbf{v}_j)$ is in W and therefore because γ is a basis of W there are scalars $a_{1,j}, a_{2,j}, \ldots, a_{m,j}$ such that

$$\mathsf{T}(\mathbf{v}_j) = a_{1,j}\mathbf{w}_1 + a_{2,j}\mathbf{w}_2 + \cdots + a_{m,j}\mathbf{w}_m$$

In other words,

$$[\mathsf{T}(v_j)]_\gamma = (a_{1,j}, a_{2,j}, \ldots, a_{m,j})$$

Substituting $\mathsf{T}(\mathbf{v}_j) = a_{1,j}\mathbf{w}_1 + a_{2,j}\mathbf{w}_2 + \cdots + a_{m,j}\mathbf{w}_m$ for each $j = 1, 2, \ldots, n$ into

$$\mathsf{T}(\mathbf{v}) = c_1\mathsf{T}(\mathbf{v}_1) + c_2\mathsf{T}(\mathbf{v}_2) + \cdots + c_n\mathsf{T}(\mathbf{v}_n)$$

and then simplifying we get

$$\mathsf{T}(\mathbf{v}) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} c_j a_{i,j}\right) w_i$$

Therefore,

$$[\mathsf{T}(\mathbf{v})]_\gamma = \mathbf{A}[\mathbf{v}]_{\mathcal{B}}$$

where $\mathbf{A}$ is the $m \times n$ matrix given by

$$\mathbf{A} = \begin{bmatrix} [\mathsf{T}(\mathbf{v}_1)]_\gamma & [\mathsf{T}(\mathbf{v}_2)]_\gamma & \cdots & [\mathsf{T}(\mathbf{v}_n)]_\gamma \end{bmatrix}$$

The matrix $\mathbf{A}$ is the matrix representation of the linear mapping T in the bases $\mathcal{B}$ and γ.

Example 18.9. Consider the vector space $\mathsf{V} = \mathbb{P}_2[t]$ of polynomial of degree no more than two and let $\mathsf{T} : \mathsf{V} \to \mathsf{V}$ be defined by

$$\mathsf{T}(\mathbf{v}(t)) = 4\mathbf{v}'(t) - 2\mathbf{v}(t)$$

It is straightforward to verify that T is a linear mapping. Let

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \{t - 1, 3 + 2t, t^2 + 1\}.$$

(a) Verify that $\mathcal{B}$ is a basis of V.

(b) Find the coordinates of $\mathbf{v}(t) = -t^2 + 3t + 1$ in the basis $\mathcal{B}$.

(c) Find the matrix representation of T in the basis $\mathcal{B}$.

www.RanaMaths.com

Solution. (a) Suppose that there are scalars c_1, c_2, c_3 such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = 0$$

Then expanding and then collecting like terms we obtain

$$c_3t^2 + (c_1 + 2c_2)t + (-c_1 + 3c_2 + c_3) = 0$$

Since the above holds for all $t \in \mathbb{R}$ we must have

$$c_3 = 0, \quad c_1 + 2c_2 = 0, \quad -c_1 + 3c_2 + c_3 = 0$$

Solving for c_1, c_2, c_3 we obtain $c_1 = 0, c_2 = 0, c_3 = 0$. Hence, the only linear combination of the vectors in $\mathcal{B}$ that produces the zero vector is the trivial linear combination. This proves by definition that $\mathcal{B}$ is linearly independent. Since we already know that $\dim(\mathbb{P}_2) = 3$ and $\mathcal{B}$ contains 3 vectors, then $\mathcal{B}$ is a basis for $\mathbb{P}_2$

(b) The coordinates of $\mathbf{v}(t) = -t^2 + 3t + 1$ are the unique scalars (c_1, c_2, c_3) such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{v}$$

In this case the linear system is

$$c_3 = -1, \quad c_1 + 2c_2 = 3, \quad -c_1 + 3c_2 + c_3 = 1$$

and solving yields $c_1 = 1, c_2 = 1$, and $c_3 = -1$. Hence,

$$[\mathbf{v}]_{\mathcal{B}} = (1, 1, -1)$$

(c) The matrix representation $\mathbf{A}$ of T is

$$\mathbf{A} = \begin{bmatrix}[\mathsf{T}(\mathbf{v}_1)]_{\mathcal{B}} & [\mathsf{T}(\mathbf{v}_2)]_{\mathcal{B}} & [\mathsf{T}(\mathbf{v}_3)]_{\mathcal{B}}\end{bmatrix}$$

Now we compute directly that

$$\mathsf{T}(\mathbf{v}_1) = -2t + 6, \quad \mathsf{T}(\mathbf{v}_2) = -4t + 2, \quad \mathsf{T}(\mathbf{v}_3) = -2t^2 + 8t - 2$$

And then one computes that

$$[\mathsf{T}(\mathbf{v}_1)]_{\mathcal{B}} = \begin{bmatrix} -18/5 \\ 4/5 \\ 0 \end{bmatrix}, \quad [\mathsf{T}(\mathbf{v}_2)]_{\mathcal{B}} = \begin{bmatrix} -6/5 \\ -2/5 \\ 0 \end{bmatrix}, \quad [\mathsf{T}(\mathbf{v}_3)]_{\mathcal{B}} = \begin{bmatrix} 24/5 \\ 8/5 \\ -2 \end{bmatrix}$$

And therefore

$$\mathbf{A} = \begin{bmatrix} -18/5 & -6/5 & 24/5 \\ 4/5 & -2/5 & 8/5 \\ 0 & 0 & -2 \end{bmatrix}$$

□

After this lecture you should know the following:

- what coordinates are (you need a basis)
- how to find coordinates relative to a basis
- the interpretation of the change-of-coordinates matrix as a mapping that transforms one set of coordinates to another

www.RanaMaths.com

www.RanaMaths.com

Lecture 19

Change of Basis

19.1 Review of Coordinate Mappings on $\mathbb{R}^n$

Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$ and let

$$\mathbf{P}_{\mathcal{B}} = [\mathbf{v}_1 \; \mathbf{v}_2 \; \cdots \; \mathbf{v}_n].$$

If $\mathbf{x} \in \mathbb{R}^n$ and $[\mathbf{x}]_{\mathcal{B}}$ is the coordinate vector of $\mathbf{x}$ in the basis $\mathcal{B}$ then

$$\mathbf{x} = \mathbf{P}_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}.$$

The components of the vector $\mathbf{x}$ are the coordinates of $\mathbf{x}$ in the standard basis $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. In other words,

$$[\mathbf{x}]_{\mathcal{E}} = \mathbf{x}.$$

Therefore,

$$[\mathbf{x}]_{\mathcal{E}} = \mathbf{P}_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}.$$

We can therefore interpret $\mathbf{P}_{\mathcal{B}}$ as the matrix mapping that maps the $\mathcal{B}$-coordinates of $\mathbf{x}$ to the $\mathcal{E}$-coordinates of $\mathbf{x}$. To make this more explicit, we sometimes use the notation

$${}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}$$

to indicate that ${}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}$ maps $\mathcal{B}$-coordinates to $\mathcal{E}$-coordinates:

$$[\mathbf{x}]_{\mathcal{E}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})[\mathbf{x}]_{\mathcal{B}}.$$

If we multiply the equation

$$[\mathbf{x}]_{\mathcal{E}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})[\mathbf{x}]_{\mathcal{B}}$$

on the left by the inverse of ${}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}$ we obtain

$$({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})^{-1}[\mathbf{x}]_{\mathcal{E}} = [\mathbf{x}]_{\mathcal{B}}$$

Hence, the matrix $({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})^{-1}$ maps standard coordinates to $\mathcal{B}$-coordinates, see Figure 19.1. It is natural then to introduce the notation

$${}_{\mathcal{B}}\mathbf{P}_{\mathcal{E}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})^{-1}$$

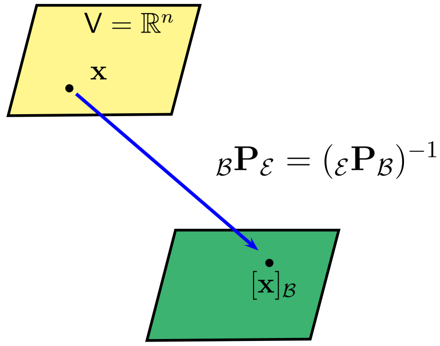


Figure 19.1: The matrix ${}_{\mathcal{B}}\mathbf{P}_{\mathcal{E}}$ maps $\mathcal{E}$ coordinates to $\mathcal{B}$ coordinates.

Example 19.1. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; \mathbf{v}_2 = \begin{bmatrix} -3 \\ 4 \\ 0 \end{bmatrix}, \; \mathbf{v}_2 = \begin{bmatrix} 3 \\ -6 \\ 3 \end{bmatrix}, \; \mathbf{x} = \begin{bmatrix} -8 \\ 2 \\ 3 \end{bmatrix}.$$

(a) Show that the set of vectors $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ forms a basis for $\mathbb{R}^n$.
(b) Find the change-of-coordinates matrix from $\mathcal{B}$ to standard coordinates.
(c) Find the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ for the given $\mathbf{x}$.

Solution. Let

$$\mathbf{P}_{\mathcal{B}} = \begin{bmatrix} 1 & -3 & 3 \\ 0 & 4 & -6 \\ 0 & 0 & 3 \end{bmatrix}$$

It is clear that $\det(\mathbf{P}_{\mathcal{B}}) = 12$, and therefore $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent. Therefore, $\mathcal{B}$ is a basis for $\mathbb{R}^n$. The matrix $\mathbf{P}_{\mathcal{B}}$ takes $\mathcal{B}$-coordinates to standard coordinates. The $\mathcal{B}$-coordinate vector $[\mathbf{x}]_{\mathcal{B}} = (c_1, c_2, c_3)$ is the unique solution to the linear system

$$\mathbf{x} = \mathbf{P}_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}$$

Solving the linear system with augmented matrix $[\mathbf{P}_{\mathcal{B}} \;\; \mathbf{x}]$ we obtain

$$[\mathbf{x}]_{\mathcal{B}} = (-5, 2, 1)$$

We verify that $[\mathbf{x}]_{\mathcal{B}} = (-5, 2, 1)$ are indeed the coordinates of $\mathbf{x} = (-8, 2, 3)$ in the basis

$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$:

$$
\begin{aligned}
(-5)\mathbf{v}_1 + (2)\mathbf{v}_2 + (1)\mathbf{v}_3 &= -5\begin{bmatrix}1\\0\\0\end{bmatrix} + 2\begin{bmatrix}-3\\4\\0\end{bmatrix} + \begin{bmatrix}3\\-6\\3\end{bmatrix}\\
&= \begin{bmatrix}-5\\0\\0\end{bmatrix} + \begin{bmatrix}-6\\8\\0\end{bmatrix} + \begin{bmatrix}3\\-6\\3\end{bmatrix}\\
&= \underbrace{\begin{bmatrix}-8\\2\\3\end{bmatrix}}_{\mathbf{x}}
\end{aligned}
$$

□

19.2 Change of Basis

We saw in the previous section that the matrix

$$_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}$$

takes as input the $\mathcal{B}$-coordinates $[\mathbf{x}]_{\mathcal{B}}$ of a vector $\mathbf{x}$ and returns the coordinates of $\mathbf{x}$ in the standard basis. We now consider the situation of dealing with two basis $\mathcal{B}$ and $\mathcal{C}$ where neither is assumed to be the standard basis $\mathcal{E}$. Hence let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and let $\mathcal{C} = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be two basis of $\mathbb{R}^n$ and let

$$_{\mathcal{E}}\mathbf{P}_{\mathcal{B}} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$$

$$_{\mathcal{E}}\mathbf{P}_{\mathcal{C}} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_n].$$

Then if $[\mathbf{x}]_{\mathcal{C}}$ is the coordinate vector of $\mathbf{x}$ in the basis $\mathcal{C}$ then

$$\mathbf{x} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})[\mathbf{x}]_{\mathcal{C}}.$$

How do we transform $\mathcal{B}$-coordinates of $\mathbf{x}$ to $\mathcal{C}$-coordinates of $\mathbf{x}$, and vice-versa? To answer this question, start from the relations

$$\mathbf{x} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})[\mathbf{x}]_{\mathcal{B}}$$

$$\mathbf{x} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})[\mathbf{x}]_{\mathcal{C}}.$$

Then

$$({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})[\mathbf{x}]_{\mathcal{C}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})[\mathbf{x}]_{\mathcal{B}}$$

and because ${}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}}$ is invertible we have that

$$[\mathbf{x}]_{\mathcal{C}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})[\mathbf{x}]_{\mathcal{B}}.$$

Hence, the matrix $({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})$ maps the $\mathcal{B}$-coordinates of $\mathbf{x}$ to the $\mathcal{C}$-coordinates of $\mathbf{x}$. For this reason, it is natural to use the notation (see Figure 19.2)

$$ {}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}). $$

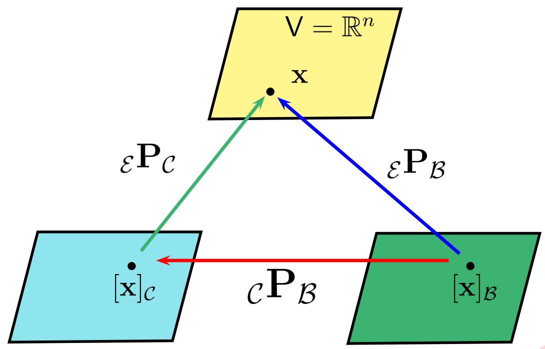


Figure 19.2: The matrix ${}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}}$ maps $\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates.

If we expand $({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})$ we obtain that

$$ ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}) = \begin{bmatrix} ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}\mathbf{v}_1 & ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}\mathbf{v}_2 & \cdots & ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}\mathbf{v}_n \end{bmatrix}. $$

Therefore, the ith column of $({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}})$, namely

$$ ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}\mathbf{v}_i, $$

is the coordinate vector of $\mathbf{v}_i$ in the basis $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$. To compute ${}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}}$ we augment ${}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}}$ and ${}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}$ and row reduce fully:

$$ \begin{bmatrix} {}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}} & {}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}} \end{bmatrix} \sim \begin{bmatrix} \mathbf{I}_n & {}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}} \end{bmatrix}. $$

Example 19.2. Let

$$ \mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \end{bmatrix} \right\}, \quad \mathcal{C} = \left\{ \begin{bmatrix} -7 \\ 9 \end{bmatrix}, \begin{bmatrix} -5 \\ 7 \end{bmatrix} \right\} $$

It can be verified that $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathcal{C} = \{\mathbf{w}_1, \mathbf{w}_2\}$ are bases for $\mathbb{R}^2$.
(a) Find the matrix the takes $\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates.
(b) Find the matrix that takes $\mathcal{C}$-coordinates to $\mathcal{B}$-coordinates.
(c) Let $\mathbf{x} = (0, -2)$. Find $[\mathbf{x}]_{\mathcal{B}}$ and $[\mathbf{x}]_{\mathcal{C}}$.

Solution. The matrix ${}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}} = [\mathbf{v}_1\ \mathbf{v}_2]$ maps $\mathcal{B}$-coordinates to standard $\mathcal{E}$-coordinates. The matrix ${}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}} = [\mathbf{w}_1\ \mathbf{w}_2]$ maps $\mathcal{C}$-coordinates to standard $\mathcal{E}$-coordinates. As we just showed, the matrix that maps $\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates is

$$ {}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}) $$

www.RanaMaths.com

It is straightforward to compute that

$$({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1} = \begin{bmatrix} -7/4 & -5/4 \\ 9/4 & 7/4 \end{bmatrix}$$

Therefore,

$${}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}} = ({}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}})^{-1}({}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}) = \begin{bmatrix} -7/4 & -5/4 \\ 9/4 & 7/4 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & -3/2 \\ -3 & 5/2 \end{bmatrix}$$

To compute ${}_{\mathcal{B}}\mathbf{P}_{\mathcal{C}}$, we can simply invert ${}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}}$. One finds that

$$({}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}})^{-1} = \begin{bmatrix} 5 & 3 \\ 6 & 4 \end{bmatrix}$$

and therefore

$${}_{\mathcal{B}}\mathbf{P}_{\mathcal{C}} = \begin{bmatrix} 5 & 3 \\ 6 & 4 \end{bmatrix}$$

Given that $\mathbf{x} = (0, -2)$, to find $[\mathbf{x}]_{\mathcal{B}}$ we must solve the linear system

$${}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$$

Row reducing the augmented matrix $[{}_{\mathcal{E}}\mathbf{P}_{\mathcal{B}}\ \mathbf{x}]$ we obtain

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Next, to find $[\mathbf{x}]_{\mathcal{C}}$ we can solve the linear system

$${}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}} = \mathbf{x}$$

Alternatively, since we now know $[\mathbf{x}]_{\mathcal{B}}$ and ${}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}}$ has been computed, to find $[\mathbf{x}]_{\mathcal{C}}$ we simply multiply ${}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}}$ by $[\mathbf{x}]_{\mathcal{B}}$:

$$[\mathbf{x}]_{\mathcal{C}} = {}_{\mathcal{C}}\mathbf{P}_{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 2 & -3/2 \\ -3 & 5/2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5/2 \\ -7/2 \end{bmatrix}$$

Let's verify that $[\mathbf{x}]_{\mathcal{C}} = \begin{bmatrix} 5/2 \\ -7/2 \end{bmatrix}$ are indeed the $\mathcal{C}$-coordinates of $\mathbf{x} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$:

$${}_{\mathcal{E}}\mathbf{P}_{\mathcal{C}}[\mathbf{x}]_{\mathcal{C}} = \begin{bmatrix} -7 & -5 \\ 9 & 7 \end{bmatrix} \begin{bmatrix} 5/2 \\ -7/2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}.$$

□

After this lecture you should know the following:

- how to compute a change of basis matrix
- and how to use the change of basis matrix to map one set of coordinates into another

www.RanaMaths.com

www.RanaMaths.com

Lecture 20

Inner Products and Orthogonality

20.1 Inner Product on $\mathbb{R}^n$

The inner product on $\mathbb{R}^n$ generalizes the notion of the **dot product** of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ that you may are already familiar with.

Definition 20.1: Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and let $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ be vectors in $\mathbb{R}^n$. The **inner product** of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \bullet \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Notice that the inner product $\mathbf{u} \bullet \mathbf{v}$ can be computed as a matrix multiplication as follows:

$$\mathbf{u} \bullet \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

The following theorem summarizes the basic algebraic properties of the inner product.

Theorem 20.2: Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be vectors in $\mathbb{R}^n$ and let α be a scalar. Then

(a) $\mathbf{u} \bullet \mathbf{v} = \mathbf{v} \bullet \mathbf{u}$

(b) $(\mathbf{u} + \mathbf{v}) \bullet \mathbf{w} = \mathbf{u} \bullet \mathbf{w} + \mathbf{v} \bullet \mathbf{w}$

(c) $(\alpha \mathbf{u}) \bullet \mathbf{v} = \alpha(\mathbf{u} \bullet \mathbf{v}) = \mathbf{u} \bullet (\alpha \mathbf{v})$

(d) $\mathbf{u} \bullet \mathbf{u} \geq 0$, and $\mathbf{u} \bullet \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

www.RanaMaths.com

Example 20.3. Let $\mathbf{u} = (2, -5, -1)$ and let $\mathbf{v} = (3, 2, -3)$. Compute $\mathbf{u} \bullet \mathbf{v}$, $\mathbf{v} \bullet \mathbf{u}$, $\mathbf{u} \bullet \mathbf{u}$, and $\mathbf{v} \bullet \mathbf{v}$.

Solution. By definition:

$$\begin{aligned}
\mathbf{u} \bullet \mathbf{v} &= (2)(3) + (-5)(2) + (1)(-3) = -1 \\
\mathbf{v} \bullet \mathbf{u} &= (3)(2) + (2)(-5) + (-3)(1) = -1 \\
\mathbf{u} \bullet \mathbf{u} &= (2)(2) + (-5)(-5) + (-1)(-1) = 30 \\
\mathbf{v} \bullet \mathbf{v} &= (3)(3) + (2)(2) + (-3)(-3) = 22.
\end{aligned}$$

□

We now define the length or norm of a vector in $\mathbb{R}^n$.

Definition 20.4: The **length** or **norm** of a vector $\mathbf{u} \in \mathbb{R}^n$ is defined as

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \bullet \mathbf{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

A vector $\mathbf{u} \in \mathbb{R}^n$ with norm 1 will be called a **unit vector**:

$$\|\mathbf{u}\| = 1.$$

Below is an important property of the inner product.

Theorem 20.5: Let $\mathbf{u} \in \mathbb{R}^n$ and let α be a scalar. Then

$$\|\alpha\mathbf{u}\| = |\alpha|\|\mathbf{u}\|.$$

Proof. We have

$$\begin{aligned}
\|\alpha\mathbf{u}\| &= \sqrt{(\alpha\mathbf{u}) \bullet (\alpha\mathbf{u})} \\
&= \sqrt{\alpha^2(\mathbf{u} \bullet \mathbf{u})} \\
&= |\alpha|\sqrt{\mathbf{u} \bullet \mathbf{u}} \\
&= |\alpha|\|\mathbf{u}\|.
\end{aligned}$$

□

By Theorem 20.5, any non-zero vector $\mathbf{u} \in \mathbb{R}^n$ can be scaled to obtain a new unit vector in the same direction as $\mathbf{u}$. Indeed, suppose that $\mathbf{u}$ is non-zero so that $\|\mathbf{u}\| \neq 0$. Define the new vector

$$\mathbf{v} = \frac{1}{\|\mathbf{u}\|}\mathbf{u}$$

www.RanaMaths.com

Notice that $\alpha = \frac{1}{\|\mathbf{u}\|}$ is just a scalar and thus $\mathbf{v}$ is a scalar multiple of $\mathbf{u}$. Then by Theorem 20.5 we have that

$$\|\mathbf{v}\| = \|\alpha\mathbf{u}\| = |\alpha| \cdot \|\mathbf{u}\| = \frac{1}{\|\mathbf{u}\|} \cdot \|\mathbf{u}\| = 1$$

and therefore $\mathbf{v}$ is a unit vector, see Figure 20.1. The process of taking a non-zero vector $\mathbf{u}$ and creating the new vector $\mathbf{v} = \frac{1}{\|\mathbf{u}\|}\mathbf{u}$ is sometimes called **normalization** of $\mathbf{u}$.

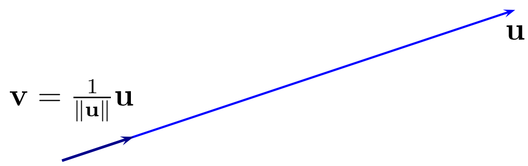


Figure 20.1: Normalizing a non-zero vector.

Example 20.6. Let $\mathbf{u} = (2, 3, 6)$. Compute $\|\mathbf{u}\|$ and find the unit vector $\mathbf{v}$ in the same direction as $\mathbf{u}$.

Solution. By definition,

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \bullet \mathbf{u}} = \sqrt{2^2 + 3^2 + 6^2} = \sqrt{49} = 7.$$

Then the unit vector that is in the same direction as $\mathbf{u}$ is

$$\mathbf{v} = \frac{1}{\|\mathbf{u}\|}\mathbf{u} = \frac{1}{7}\begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 3/7 \\ 6/7 \end{bmatrix}$$

Verify that $\|\mathbf{v}\| = 1$:

$$\|\mathbf{v}\| = \sqrt{(2/7)^2 + (3/7)^2 + (6/7)^2} = \sqrt{4/49 + 9/49 + 36/49} = \sqrt{49/49} = \sqrt{1} = 1.$$

□

Now that we have the definition of the length of a vector, we can define the notion of distance between two vectors.

Definition 20.7: Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. The **distance between u and v** is the length of the vector $\mathbf{u} - \mathbf{v}$. We will denote the distance between $\mathbf{u}$ and $\mathbf{v}$ by $\mathrm{d}(\mathbf{u}, \mathbf{v})$. In other words,

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Example 20.8. Find the distance between $\mathbf{u} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 7 \\ -9 \end{bmatrix}$.

Solution. We compute:

$$\mathrm{d}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(3-7)^2 + (-2+9)^2} = \sqrt{65}.$$

□

20.2 Orthogonality

In the context of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, orthogonality is synonymous with perpendicularity. Below is the general definition.

Definition 20.9: Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are said to be **orthogonal** if $\mathbf{u} \bullet \mathbf{v} = 0$.

In $\mathbb{R}^2$ and $\mathbb{R}^3$, the notion of orthogonality should be familiar to you. In fact, using the Law of Cosines in $\mathbb{R}^2$ or $\mathbb{R}^3$, one can prove that

$$\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cos(\theta) \tag{20.1}$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. If $\theta = \frac{\pi}{2}$ then clearly $\mathbf{u} \bullet \mathbf{v} = 0$. In higher dimensions, i.e., $n \geq 4$, we can use equation (20.1) to *define* the angle between vectors $\mathbf{u}$ and $\mathbf{v}$. In other words, the **angle** between any two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is define to be

$$\theta = \arccos\left(\frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}\right).$$

The general notion of orthogonality in $\mathbb{R}^n$ leads to the following theorem from grade school.

Theorem 20.10: (Pythagorean Theorem) Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Solution. First recall that $\|\mathbf{u}+\mathbf{v}\| = \sqrt{(\mathbf{u}+\mathbf{v}) \bullet (\mathbf{u}+\mathbf{v})}$ and therefore

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v}) \bullet (\mathbf{u}+\mathbf{v}) \\
&= \mathbf{u} \bullet \mathbf{u} + \mathbf{u} \bullet \mathbf{v} + \mathbf{v} \bullet \mathbf{u} + \mathbf{v} \bullet \mathbf{v} \\
&= \|\mathbf{u}\|^2 + 2(\mathbf{u} \bullet \mathbf{v}) + \|\mathbf{v}\|^2.
\end{aligned}$$

Therefore, $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $\mathbf{u} \bullet \mathbf{v} = 0$. □

We now introduce orthogonal sets.

Definition 20.11: A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is said to be an **orthogonal set** if any pair of distinct vectors $\mathbf{u}_i, \mathbf{u}_j$ are orthogonal, that is, $\mathbf{u}_i \bullet \mathbf{u}_j = 0$ whenever $i \neq j$.

In the following theorem we prove that orthogonal sets are linearly independent.

www.RanaMaths.com

Theorem 20.12: Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be an orthogonal set of non-zero vectors in $\mathbb{R}^n$. Then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is linearly independent. In particular, if $p = n$ then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is basis for $\mathbb{R}^n$.

Solution. Suppose that there are scalars $c_1, c_2, \ldots, c_p$ such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p = \mathbf{0}.$$

Take the inner product of $\mathbf{u}_1$ with both sides of the above equation:

$$c_1(\mathbf{u}_1 \bullet \mathbf{u}_1) + c_2(\mathbf{u}_2 \bullet \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p \bullet \mathbf{u}_1) = \mathbf{0} \bullet \mathbf{u}_1.$$

Since the set is orthogonal, the left-hand side of the last equation simplifies to $c_1(\mathbf{u}_1 \bullet \mathbf{u}_1)$. The right-hand side simplifies to $\mathbf{0}$. Hence,

$$c_1(\mathbf{u}_1 \bullet \mathbf{u}_1) = \mathbf{0}.$$

But $\mathbf{u}_1 \bullet \mathbf{u}_1 = \|\mathbf{u}_1\|^2$ is not zero and therefore the only way that $c_1(\mathbf{u}_1 \bullet \mathbf{u}_2) = \mathbf{0}$ is if $c_1 = 0$. Repeat the above steps using $\mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_p$ and conclude that $c_2 = 0$, $c_3 = 0, \ldots, c_p = 0$. Therefore, $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is linearly independent. If $p = n$, then the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is automatically a basis for $\mathbb{R}^n$. □

Example 20.13. Is the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ an orthogonal set?

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix}$$

Solution. Compute

$$\begin{aligned} \mathbf{u}_1 \bullet \mathbf{u}_2 &= (1)(0) + (-2)(1) + (1)(2) = 0 \\ \mathbf{u}_1 \bullet \mathbf{u}_3 &= (1)(-5) + (-2)(-2) + (1)(1) = 0 \\ \mathbf{u}_2 \bullet \mathbf{u}_3 &= (0)(-5) + (1)(-2) + (2)(1) = 0 \end{aligned}$$

Therefore, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set. By Theorem 20.12, the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly independent. To verify linear independence, we computed that $\det(\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3\end{bmatrix}) = 30$, which is non-zero. □

www.RanaMaths.com

We now introduce orthonormal sets.

Definition 20.14: A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ is said to be an **orthonormal set** if it is an orthogonal set and if each vector $\mathbf{u}_i$ in the set is a unit vector.

Consider the previous orthogonal set in $\mathbb{R}^3$:

$$\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -5 \\ -2 \\ 1 \end{bmatrix} \right\}.$$

It is not an orthonormal set because none of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are unit vectors. Explicitly, $\|\mathbf{u}_1\| = \sqrt{6}$, $\|\mathbf{u}_2\| = \sqrt{5}$, and $\|\mathbf{u}_3\| = \sqrt{30}$. However, from an **orthogonal set** we can create an **orthonormal set** by normalizing each vector. Hence, the set

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} 0 \\ 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}, \begin{bmatrix} -5/\sqrt{30} \\ -2/\sqrt{30} \\ 1/\sqrt{30} \end{bmatrix} \right\}$$

is an orthonormal set.

20.3 Coordinates in an Orthonormal Basis

As we will see in this section, a basis $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of $\mathbb{R}^n$ that is also an orthonormal set is highly desirable when performing computations with coordinates. To see why, let $\mathbf{x}$ be any vector in $\mathbb{R}^n$ and suppose we want to find the coordinates of $\mathbf{x}$ in the basis $\mathcal{B}$, that is we seek to find $[\mathbf{x}]_{\mathcal{B}} = (c_1, c_2, \ldots, c_n)$. By definition, the coordinates $c_1, c_2, \ldots, c_n$ satisfy the equation

$$\mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n.$$

Taking the inner product of $\mathbf{u}_1$ with both sides of the above equation and using the fact that $\mathbf{u}_1 \bullet \mathbf{u}_2 = 0$, $\mathbf{u}_1 \bullet \mathbf{u}_3 = 0$, and $\mathbf{u}_1 \bullet \mathbf{u}_n = 0$, we obtain

$$\mathbf{u}_1 \bullet \mathbf{x} = c_1(\mathbf{u}_1 \bullet \mathbf{u}_1) = c_1(1) = c_1$$

where we also used the fact that $\mathbf{u}_i$ is a unit vector. Thus, $c_1 = \mathbf{u}_1 \bullet \mathbf{x}$! Repeating this procedure with $\mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n$ we obtain the remaining coefficients $c_2, \ldots, c_n$:

$$\begin{aligned} c_2 &= \mathbf{u}_2 \bullet \mathbf{x} \\ c_3 &= \mathbf{u}_3 \bullet \mathbf{x} \\ \vdots &= \;\vdots \\ c_n &= \mathbf{u}_n \bullet \mathbf{x}. \end{aligned}$$

Our previous computation proves the following theorem.

www.RanaMaths.com

Theorem 20.15: Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for $\mathbb{R}^n$. The coordinate vector of $\mathbf{x}$ in the basis $\mathcal{B}$ is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \mathbf{u}_1 \bullet \mathbf{x} \\ \mathbf{u}_2 \bullet \mathbf{x} \\ \vdots \\ \mathbf{u}_n \bullet \mathbf{x} \end{bmatrix}.$$

Hence, computing coordinates with respect to an orthonormal basis can be done without performing any row operations and all we need to do is compute inner products! We make the important observation that an alternate expression for $[\mathbf{x}]_{\mathcal{B}}$ is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \mathbf{u}_1 \bullet \mathbf{x} \\ \mathbf{u}_2 \bullet \mathbf{x} \\ \vdots \\ \mathbf{u}_n \bullet \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \mathbf{x} = \mathbf{U}^T\mathbf{x}$$

where $\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n]$. On the other hand, recall that by definition $[\mathbf{x}]_{\mathcal{B}}$ satisfies $\mathbf{U}[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}$, and therefore $[\mathbf{x}]_{\mathcal{B}} = \mathbf{U}^{-1}\mathbf{x}$. If we compare the two identities

$$[\mathbf{x}]_{\mathcal{B}} = \mathbf{U}^{-1}\mathbf{x} \quad \text{and} \quad [\mathbf{x}]_{\mathcal{B}} = \mathbf{U}^T\mathbf{x}$$

we suspect then that $\mathbf{U}^{-1} = \mathbf{U}^T$. This is indeed the case. To see this, let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ and put

$$\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n].$$

Consider the matrix product $\mathbf{U}^T\mathbf{U}$, and recalling that $\mathbf{u}_i \bullet \mathbf{u}_j = \mathbf{u}_i^T\mathbf{u}_j$, we obtain

$$\mathbf{U}^T\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \cdots & \mathbf{u}_1^T\mathbf{u}_n \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \cdots & \mathbf{u}_2^T\mathbf{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_n^T\mathbf{u}_1 & \mathbf{u}_n^T\mathbf{u}_2 & \cdots & \mathbf{u}_n^T\mathbf{u}_n \end{bmatrix}$$

$$= \mathbf{I}_n.$$

Therefore,

$$\mathbf{U}^{-1} = \mathbf{U}^T.$$

A matrix $\mathbf{U} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_n$$

is called a **orthogonal matrix**. Hence, if $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal set then the matrix

$$\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}$$

is an orthogonal matrix.

Example 20.16. Consider the vectors

$$\mathbf{v}_1 = \begin{bmatrix}1\\0\\1\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}-1\\4\\1\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}2\\1\\-2\end{bmatrix}, \mathbf{x} = \begin{bmatrix}1\\2\\-1\end{bmatrix}.$$

(a) Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for $\mathbb{R}^3$.
(b) Then, if necessary, normalize the basis vectors $\mathbf{v}_i$ to obtain an orthonormal basis $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for $\mathbb{R}^3$.
(c) For the given $\mathbf{x}$ find $[\mathbf{x}]_{\mathcal{B}}$.

Solution. (a) We compute that $\mathbf{v}_1 \bullet \mathbf{v}_2 = 0$, $\mathbf{v}_1 \bullet \mathbf{v}_3 = 0$, and $\mathbf{v}_2 \bullet \mathbf{v}_3 = 0$, and thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set. Since orthogonal sets are linearly independent and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ consists of three vectors then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is basis for $\mathbb{R}^3$.

(b) We compute that $\|\mathbf{v}_1\| = \sqrt{2}$, $\|\mathbf{v}_2\| = \sqrt{18}$, and $\|\mathbf{v}_3\| = 3$. Then let

$$\mathbf{u}_1 = \begin{bmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix}-1/\sqrt{18}\\4/\sqrt{18}\\1/\sqrt{18}\end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix}2/3\\1/3\\-2/3\end{bmatrix}$$

Then $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is now an orthonormal set and thus since $\mathcal{B}$ consists of three vectors then $\mathcal{B}$ is an orthonormal basis of $\mathbb{R}^3$.

(c) Finally, computing coordinates in an orthonormal basis is easy:

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix}\mathbf{u}_1 \bullet \mathbf{x}\\\mathbf{u}_2 \bullet \mathbf{x}\\\mathbf{u}_3 \bullet \mathbf{x}\end{bmatrix} = \begin{bmatrix}0\\2/\sqrt{18}\\5/3\end{bmatrix}$$

□

Example 20.17. The standard unit basis

$$\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

www.RanaMaths.com

in $\mathbb{R}^3$ is an orthonormal basis. Given any $\mathbf{x} = (x_1, x_2, x_3)$, we have $[\mathbf{x}]_{\mathcal{E}} = \mathbf{x}$. On the other hand, clearly

$$x_1 = \mathbf{x} \bullet \mathbf{e}_1$$
$$x_2 = \mathbf{x} \bullet \mathbf{e}_2$$
$$x_3 = \mathbf{x} \bullet \mathbf{e}_3$$

Example 20.18. (Orthogonal Complements) Let W be a subspace of $\mathbb{R}^n$. The **orthogonal complement** of W, which we denote by $\mathsf{W}^\perp$, consists of the vectors in $\mathbb{R}^n$ that are orthogonal to every vector in W. Using set notation:

$$\mathsf{W}^\perp = \{\mathbf{u} \in \mathbb{R}^n \,:\, \mathbf{u} \bullet \mathbf{w} = 0 \text{ for every } \mathbf{w} \in \mathsf{W}\}.$$

(a) Show that $\mathsf{W}^\perp$ is a subspace.
(b) Let $\mathbf{w}_1 = (0, 1, 1, 0)$, let $\mathbf{w}_2 = (1, 0, -1, 0)$, and let $\mathsf{W} = \text{span}\{\mathbf{w}_1, \mathbf{w}_2\}$. Find a basis for $\mathsf{W}^\perp$.

Solution. (a) The vector $\mathbf{0}$ is orthogonal to every vector in $\mathbb{R}^n$ and therefore it is certainly orthogonal to every vector in W. Thus, $\mathbf{0} \in \mathsf{W}^\perp$. Now suppose that $\mathbf{u}_1, \mathbf{u}_2$ are two vectors in $\mathsf{W}^\perp$. Then for any vector $\mathbf{w} \in \mathsf{W}$ it holds that

$$(\mathbf{u}_1 + \mathbf{u}_2) \bullet \mathbf{w} = \mathbf{u}_1 \bullet \mathbf{w} + \mathbf{u}_2 \bullet \mathbf{w} = 0 + 0 = 0.$$

Therefore, $\mathbf{u}_1 + \mathbf{u}_2$ is also orthogonal to $\mathbf{w}$ and since $\mathbf{w}$ is an arbitrary vector in W then $(\mathbf{u}_1 + \mathbf{u}_2) \in \mathsf{W}^\perp$. Lastly, let α be any scalar and let $\mathbf{u} \in \mathsf{W}^\perp$. Then for any vector $\mathbf{w}$ in W we have that

$$(\alpha\mathbf{u}) \bullet \mathbf{w} = \alpha(\mathbf{u} \bullet \mathbf{w}) = \alpha \cdot 0 = 0.$$

Therefore, $\alpha\mathbf{u}$ is orthogonal to $\mathbf{w}$ and since $\mathbf{w}$ is an arbitrary vector in W then $(\alpha\mathbf{u}) \in \mathsf{W}^\perp$. This proves that $\mathsf{W}^\perp$ is a subspace of $\mathbb{R}^n$.

(b) A vector $\mathbf{u} = (u_1, u_2, u_3, u_3)$ is in $\mathsf{W}^\perp$ if $\mathbf{u} \bullet \mathbf{w}_1 = 0$ and $\mathbf{u} \bullet \mathbf{w}_2 = 0$. In other words, if

$$u_2 + u_3 = 0$$
$$u_1 - u_3 = 0$$

This is a linear system for the unknowns u_1, u_2, u_3, u_4. The general solution to the linear system is

$$\mathbf{u} = t \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}.$$

Therefore, a basis for $\mathsf{W}^\perp$ is $\{(1, 0, 1, 0), (0, 1, -1, 0)\}$.

□

After this lecture you should know the following:

www.RanaMaths.com

- how to compute inner products, norms, and distances
- how to normalize vectors to unit length
- what orthogonality is and how to check for it
- what an orthogonal and orthonormal basis is
- the advantages of working with orthonormal basis when computing coordinate vectors

www.RanaMaths.com

Lecture 21

Eigenvalues and Eigenvectors

21.1 Eigenvectors and Eigenvalues

An $n \times n$ matrix $\mathbf{A}$ can be thought of as the linear mapping that takes any arbitrary vector $\mathbf{x} \in \mathbb{R}^n$ and outputs a new vector $\mathbf{Ax}$. In some cases, the new output vector $\mathbf{Ax}$ is simply a scalar multiple of the input vector $\mathbf{x}$, that is, there exists a scalar λ such that $\mathbf{Ax} = \lambda\mathbf{x}$. This case is so important that we make the following definition.

Definition 21.1: Let $\mathbf{A}$ be a $n \times n$ matrix and let $\mathbf{v}$ be a non-zero vector. If $\mathbf{Av} = \lambda\mathbf{v}$ for some scalar λ then we call the vector $\mathbf{v}$ an **eigenvector** of $\mathbf{A}$ and we call the scalar λ an **eigenvalue** of $\mathbf{A}$ corresponding to $\mathbf{v}$.

Hence, an eigenvector $\mathbf{v}$ of $\mathbf{A}$ is simply scaled by a scalar λ under multiplication by $\mathbf{A}$. Eigenvectors are by definition nonzero vectors because $\mathbf{A0}$ is clearly a scalar multiple of $\mathbf{0}$ and then it is not clear what that the corresponding eigenvalue should be.

Example 21.2. Determine if the given vectors $\mathbf{v}$ and $\mathbf{u}$ are eigenvectors of $\mathbf{A}$? If yes, find the eigenvalue of $\mathbf{A}$ associated to the eigenvector.

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

Solution. Compute

$$\begin{aligned} \mathbf{Av} = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} -6 \\ 0 \\ 2 \end{bmatrix} \\ &= 2 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \\ &= 2\mathbf{v} \end{aligned}$$

www.RanaMaths.com

Hence, $\mathbf{A}\mathbf{v} = 2\mathbf{v}$ and thus $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ with corresponding eigenvalue $\lambda = 2$. On the other hand,

$$\mathbf{A}\mathbf{u} = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 4 \end{bmatrix}.$$

There is no scalar λ such that

$$\begin{bmatrix} 0 \\ 6 \\ 4 \end{bmatrix} = \lambda \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

Therefore, $\mathbf{u}$ is not an eigenvector of $\mathbf{A}$. □

Example 21.3. Is $\mathbf{v}$ an eigenvector of $\mathbf{A}$? If yes, find the eigenvalue of $\mathbf{A}$ associated to v:

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -4 & 2 & 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Solution. We compute

$$\mathbf{A}\mathbf{v} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}.$$

Hence, if $\lambda = 0$ then $\lambda\mathbf{v} = \mathbf{0}$ and thus $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Therefore, $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ with corresponding eigenvalue $\lambda = 0$. □

How does one find the eigenvectors/eigenvalues of a matrix $\mathbf{A}$? The general procedure is to first find the eigenvalues of $\mathbf{A}$ and then for each eigenvalue find the corresponding eigenvectors. In this section, however, we will instead suppose that we have already found the eigenvalues of $\mathbf{A}$ and concern ourselves with finding the associated eigenvectors. Suppose then that λ is known to be an eigenvalue of $\mathbf{A}$. How do we find an eigenvector $\mathbf{v}$ corresponding to the eigenvalue λ? To answer this question, we note that if $\mathbf{v}$ is to be an eigenvector of $\mathbf{A}$ with eigenvalue λ then $\mathbf{v}$ must satisfy the equation

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}.$$

We can rewrite this equation as

$$\mathbf{A}\mathbf{v} - \lambda\mathbf{v} = \mathbf{0}$$

which, after using the distributive property of matrix multiplication, is equivalent to

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}.$$

The last equation says that if $\mathbf{v}$ is to be an eigenvector of $\mathbf{A}$ with eigenvalue λ then $\mathbf{v}$ must be in the null space of $\mathbf{A} - \lambda\mathbf{I}$:

$$\mathbf{v} \in \text{Null}(\mathbf{A} - \lambda\mathbf{I}).$$

www.RanaMaths.com

In summary, if λ is known to be an eigenvalue of $\mathbf{A}$, then to find the eigenvectors corresponding to λ we must solve the homogeneous system

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}.$$

Recall that the null space of any matrix is a subspace and for this reason we call the subspace $\text{Null}(\mathbf{A} - \lambda\mathbf{I})$ the **eigenspace** of $\mathbf{A}$ corresponding to λ.

Example 21.4. It is known that $\lambda = 4$ is an eigenvalue of

$$\mathbf{A} = \begin{bmatrix} -4 & 6 & 3 \\ 1 & 7 & 9 \\ 8 & -6 & 1 \end{bmatrix}.$$

Find a basis for the eigenspace of $\mathbf{A}$ corresponding to $\lambda = 4$.

Solution. First compute

$$\mathbf{A} - 4\mathbf{I} = \begin{bmatrix} -4 & 6 & 3 \\ 1 & 7 & 9 \\ 8 & -6 & 1 \end{bmatrix} - \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} -8 & 6 & 3 \\ 1 & 3 & 9 \\ 8 & -6 & -3 \end{bmatrix}$$

Find a basis for the null space of $\mathbf{A} - 4\mathbf{I}$:

$$\begin{bmatrix} -8 & 6 & 3 \\ 1 & 3 & 9 \\ 8 & -6 & -3 \end{bmatrix} \xrightarrow{R_1 \updownarrow R_2} \begin{bmatrix} 1 & 3 & 9 \\ -8 & 6 & 3 \\ 8 & -6 & -3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 9 \\ -8 & 6 & 3 \\ 8 & -6 & -3 \end{bmatrix} \xrightarrow[-8R_1+R_3]{8R_1+R_2} \begin{bmatrix} 1 & 3 & 9 \\ 0 & 30 & 75 \\ 0 & -30 & -75 \end{bmatrix}$$

Finally,

$$\begin{bmatrix} 1 & 3 & 9 \\ 0 & 30 & 75 \\ 0 & -30 & -75 \end{bmatrix} \xrightarrow{R_2+R_3} \begin{bmatrix} 1 & 3 & 9 \\ 0 & 30 & 75 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, the general solution to the homogenous system $(\mathbf{A} - 4\mathbf{I})\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = t \begin{bmatrix} -3/2 \\ -5/2 \\ 1 \end{bmatrix}$$

where t is an arbitrary scalar. Therefore, the eigenspace of $\mathbf{A}$ corresponding to $\lambda = 4$ is

$$\text{span}\left\{ \begin{bmatrix} -3/2 \\ -5/2 \\ 1 \end{bmatrix} \right\} = \text{span}\left\{ \begin{bmatrix} -3 \\ -5 \\ 2 \end{bmatrix} \right\} = \text{span}\{\mathbf{v}\}$$

and $\{\mathbf{v}\}$ is a basis for the eigenspace. The vector $\mathbf{v}$ is of course an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda = 4$ and also (of course) any multiple of $\mathbf{v}$ is also eigenvector of $\mathbf{A}$ with $\lambda = 4$. □

www.RanaMaths.com

Example 21.5. It is known that $\lambda = 3$ is an eigenvalue of

$$\mathbf{A} = \begin{bmatrix} 11 & -4 & -8 \\ 4 & 1 & -4 \\ 8 & -4 & -5 \end{bmatrix}.$$

Find the eigenspace of $\mathbf{A}$ corresponding to $\lambda = 3$.

Solution. First compute

$$\mathbf{A} - 3\mathbf{I} = \begin{bmatrix} 11 & -4 & -8 \\ 4 & 1 & -4 \\ 8 & -4 & -5 \end{bmatrix} - \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 8 & -4 & -8 \\ 4 & -2 & -4 \\ 8 & -4 & -8 \end{bmatrix}$$

Now find the null space of $\mathbf{A} - 3\mathbf{I}$:

$$\begin{bmatrix} 8 & -4 & -8 \\ 4 & -2 & -4 \\ 8 & -4 & -8 \end{bmatrix} \xrightarrow{R_1 \updownarrow R_2} \begin{bmatrix} 4 & -2 & -4 \\ 8 & -4 & -8 \\ 8 & -4 & -8 \end{bmatrix}$$

$$\begin{bmatrix} 4 & -2 & -4 \\ 8 & -4 & -8 \\ 8 & -4 & -8 \end{bmatrix} \xrightarrow[-2R_1+R_3]{-2R_1+R_2} \begin{bmatrix} 4 & -2 & -4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, any vector in the null space of

$$\mathbf{A} - 3\mathbf{I} = \begin{bmatrix} 4 & -2 & -4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

can be written as

$$\mathbf{x} = t_1 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + t_2 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

Therefore, the eigenspace of $\mathbf{A}$ corresponding to $\lambda = 3$ is

$$\text{Null}(\mathbf{A} - 3\mathbf{I}) = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right\}.$$

The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are two linearly independent eigenvectors of $\mathbf{A}$ with eigenvalue $\lambda = 3$. Therefore $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for the eigenspace of $\mathbf{A}$ with eigenvalue $\lambda = 3$. You can verify that $\mathbf{A}\mathbf{v}_1 = 3\mathbf{v}_1$ and $\mathbf{A}\mathbf{v}_2 = 3\mathbf{v}_2$. □

As shown in the last example, there may exist more than one linearly independent eigenvector of $\mathbf{A}$ corresponding to the same eigenvalue, in other words, it is possible that the dimension of the eigenspace $\text{Null}(\mathbf{A} - \lambda\mathbf{I})$ is greater than one. What can be said about the eigenvectors of $\mathbf{A}$ corresponding to different eigenvalues?

Theorem 21.6: Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be eigenvectors of $\mathbf{A}$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ of $\mathbf{A}$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly independent set.

Solution. Suppose by contradiction that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent and $\{\lambda_1, \ldots, \lambda_k\}$ are distinct. Then, one of the eigenvectors $\mathbf{v}_{p+1}$ that is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$, and $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent:

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p. \tag{21.1}$$

Applying $\mathbf{A}$ to both sides we obtain

$$\mathbf{A}\mathbf{v}_{p+1} = c_1\mathbf{A}\mathbf{v}_1 + c_2\mathbf{A}\mathbf{v}_2 + \cdots + c_p\mathbf{A}\mathbf{v}_p$$

and since $\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i$ we can simplify this to

$$\lambda_{p+1}\mathbf{v}_{p+1} = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_p\lambda_p\mathbf{v}_p. \tag{21.2}$$

On the other hand, multiply (21.1) by λ_{p+1}:

$$\lambda_{p+1}\mathbf{v}_{p+1} = c_1\lambda_{p+1}\mathbf{v}_1 + c_2\lambda_{p+1}\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p\lambda_{p+1}. \tag{21.3}$$

Now subtract equations (21.2) and (21.3):

$$\mathbf{0} = c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + c_2(\lambda_2 - \lambda_{p+1})\mathbf{v}_2 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p.$$

Now $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent and thus $c_i(\lambda_i - \lambda_{p+1}) = 0$. But the eigenvalues $\{\lambda_1, \ldots, \lambda_k\}$ are all distinct and so must have $c_1 = c_2 = \cdots = c_p = 0$. But from (21.1) this implies that $\mathbf{v}_{p+1} = \mathbf{0}$, which is a contradiction because eigenvectors are by definition non-zero. This proves that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent set. $\square$

Example 21.7. It is known that $\lambda_1 = 1$ and $\lambda_2 = -1$ are eigenvalues of

$$\mathbf{A} = \begin{bmatrix} -4 & 6 & 3 \\ 1 & 7 & 9 \\ 8 & -6 & 1 \end{bmatrix}.$$

Find bases for the eigenspaces corresponding to λ_1 and λ_2 and show that any two vectors from these distinct eigenspaces are linearly independent.

Solution. Compute

$$\mathbf{A} - \lambda_1\mathbf{I} = \begin{bmatrix} -5 & 6 & 3 \\ 1 & 6 & 9 \\ 8 & -6 & 0 \end{bmatrix}$$

and one finds that

$$(\mathbf{A} - \lambda_1\mathbf{I}) = \text{span}\left\{ \begin{bmatrix} -3 \\ -4 \\ 3 \end{bmatrix} \right\}$$

www.RanaMaths.com

Hence, $\mathbf{v}_1 = (-3, -4, 3)$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda_1 = 1$, and $\{\mathbf{v}_1\}$ forms a basis for the corresponding eigenspace. Next, compute

$$\mathbf{A} - \lambda_2 \mathbf{I} = \begin{bmatrix} -4 & 6 & 3 \\ 1 & 7 & 9 \\ 8 & -6 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 6 & 3 \\ 1 & 8 & 9 \\ 8 & -6 & 2 \end{bmatrix}$$

and one finds that

$$\mathbf{A} - \lambda_2 \mathbf{I} = \text{span}\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

Hence, $\mathbf{v}_2 = (-1, -1, 1)$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda_2 = -1$, and $\{\mathbf{v}_2\}$ forms a basis for the corresponding eigenspace. Now verify that $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent:

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} -3 & -1 \\ -4 & -1 \\ 3 & 1 \end{bmatrix} \xrightarrow{R_1 + R_3} \begin{bmatrix} -3 & -1 \\ -4 & -1 \\ 0 & 0 \end{bmatrix}$$

The last matrix has rank $r = 2$, and thus $\mathbf{v}_1, \mathbf{v}_2$ are indeed linearly independent. □

21.2 When $\lambda = 0$ is an eigenvalue

What can we say about $\mathbf{A}$ if $\lambda = 0$ is an eigenvalue of $\mathbf{A}$? Suppose then that $\mathbf{A}$ has eigenvalue $\lambda = 0$. Then by definition, there exists a non-zero vector $\mathbf{v}$ such that

$$\mathbf{A}\mathbf{v} = 0 \cdot \mathbf{v} = \mathbf{0}.$$

In other words, $\mathbf{v}$ is in the null space of $\mathbf{A}$. Thus, $\mathbf{A}$ is not invertible (Why?).

Theorem 21.8: The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible if and only if $\lambda = 0$ is not an eigenvalue of $\mathbf{A}$.

In fact, later we will see that $\det(\mathbf{A})$ is the product of its eigenvalues.

After this lecture you should know the following:

- what eigenvalues are
- what eigenvectors are and how to find them when eigenvalues are known
- the behavior of a discrete dynamical system when the initial condition is set to an eigenvector of the system matrix

Lecture 22

The Characteristic Polynomial

22.1 The Characteristic Polynomial of a Matrix

Recall that a number λ is an eigenvalue of $\mathbf{A} \in \mathbb{R}^{n\times n}$ if there exists a non-zero vector $\mathbf{v}$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

or equivalently if $\mathbf{v} \in \text{Null}(\mathbf{A} - \lambda\mathbf{I})$. In other words, λ is an eigenvalue of $\mathbf{A}$ if and only if the subspace $\text{Null}(\mathbf{A} - \lambda\mathbf{I})$ contains a vector other than the zero vector. We know that any matrix $\mathbf{M}$ has a non-trivial null space if and only if $\mathbf{M}$ is non-invertible if and only if $\det(\mathbf{M}) = 0$. Hence, λ is an eigenvalue of $\mathbf{A}$ if and only if λ satisfies $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Let's compute the expression $\det(\mathbf{A} - \lambda\mathbf{I})$ for a generic 2×2 matrix:

$$\begin{aligned}\det(\mathbf{A} - \lambda\mathbf{I}) &= \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} \\ &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{22} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{22}.\end{aligned}$$

Thus, if $\mathbf{A}$ is 2×2 then

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{22}$$

is a polynomial in the variable λ of degree $n = 2$. This motivates the following definition.

Definition 22.1: Let $\mathbf{A}$ be a $n \times n$ matrix. The polynomial

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$$

is called the **characteristic polynomial** of $\mathbf{A}$.

www.RanaMaths.com

In summary, to find the eigenvalues of $\mathbf{A}$ we must find the roots of the characteristic polynomial:

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}).$$

The following theorem asserts that what we observed for the case $n = 2$ is indeed true for all n.

Theorem 22.2: The characteristic polynomial $p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$ of a $n \times n$ matrix $\mathbf{A}$ is an nth degree polynomial.

Solution. Recall that for the case $n = 2$ we computed that

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{22}.$$

Therefore, the claim holds for $n = 2$. By induction, suppose that the claims hold for $n \geq 2$. If $\mathbf{A}$ is a $(n+1) \times (n+1)$ matrix then expanding $\det(\mathbf{A} - \lambda\mathbf{I})$ along the first row:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (a_{11} - \lambda)\det(\mathbf{A}_{11} - \lambda\mathbf{I}) + \sum_{k=2}^{n} (-1)^{1+k} a_{1k} \det(\mathbf{A}_{1k} - \lambda\mathbf{I}).$$

By induction, each of $\det(\mathbf{A}_{1k} - \lambda\mathbf{I})$ is a nth degree polynomial. Hence, $(a_{11} - \lambda)\det(\mathbf{A}_{11} - \lambda\mathbf{I})$ is a $(n+1)$th degree polynomial. This ends the proof. □

Example 22.3. Find the characteristic polynomial of

$$\mathbf{A} = \begin{bmatrix} -2 & 4 \\ -6 & 8 \end{bmatrix}.$$

What are the eigenvalues of $\mathbf{A}$?

Solution. Compute

$$\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} -2 & 4 \\ -6 & 8 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} -2-\lambda & 4 \\ -6 & 8-\lambda \end{bmatrix}.$$

Therefore,

$$\begin{aligned} p(\lambda) &= \det(\mathbf{A} - \lambda\mathbf{I}) \\ &= \begin{vmatrix} -2-\lambda & 4 \\ -6 & 8-\lambda \end{vmatrix} \\ &= (-2-\lambda)(8-\lambda) + 24 \\ &= \lambda^2 - 6\lambda + 8 \\ &= (\lambda - 4)(\lambda - 2) \end{aligned}$$

The roots of $p(\lambda)$ are clearly $\lambda_1 = 4$ and $\lambda_2 = 2$. Therefore, the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 4$ and $\lambda_2 = 2$. □

www.RanaMaths.com

Example 22.4. Find the eigenvalues of

$$\mathbf{A} = \begin{bmatrix} -4 & -6 & -7 \\ 3 & 5 & 3 \\ 0 & 0 & 3 \end{bmatrix}.$$

Solution. Compute

$$\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} -4 & -6 & -7 \\ 3 & 5 & 3 \\ 0 & 0 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$

$$= \begin{bmatrix} -4-\lambda & -6 & -7 \\ 3 & 5-\lambda & 3 \\ 0 & 0 & 3-\lambda \end{bmatrix}$$

Then

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= (-4-\lambda)\begin{vmatrix} 5-\lambda & 3 \\ -\lambda & 3-\lambda \end{vmatrix} - 3\begin{vmatrix} -6 & -7 \\ -\lambda & 3-\lambda \end{vmatrix} \\ &= (-4-\lambda)[(3-\lambda)(5-\lambda)+3\lambda] - 3[-6(3-\lambda)-7\lambda] \\ &= \lambda^3 - 4\lambda^2 + \lambda + 6 \end{aligned}$$

Factor the characteristic polynomial:

$$p(\lambda) = \lambda^3 - 4\lambda^2 + \lambda + 6 = (\lambda-2)(\lambda-3)(\lambda+1)$$

Therefore, the eigenvalues of **A** are

$$\lambda_1 = 2, \quad \lambda_2 = 3, \quad \lambda_3 = -1.$$

□

Now that we know how to find eigenvalues, we can combine our work from the previous lecture to find both the eigenvalues and eigenvectors of a given matrix **A**.

Example 22.5. For each eigenvalue of **A** from Example 22.4, find a basis for the corresponding eigenspace.

Solution. Start with $\lambda_1 = 2$:

$$\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} -6 & -6 & -7 \\ 3 & 3 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

After basic row reduction and back substitution, one finds that the null space of $\mathbf{A} - 2\mathbf{I}$ is spanned by

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}.$$

www.RanaMaths.com

Therefore, $\mathbf{v}_1$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ_1. For $\lambda_2 = 3$:

$$\mathbf{A} - 3\mathbf{I} = \begin{bmatrix} -7 & -6 & -7 \\ 3 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

The null space of $\mathbf{A} - 3\mathbf{I}$ is spanned by

$$\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

and therefore $\mathbf{v}_2$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ_2. Finally, for $\lambda_3 = -1$ we compute

$$\mathbf{A} - \lambda_3\mathbf{I} = \begin{bmatrix} -3 & -6 & -7 \\ 3 & 6 & 3 \\ 0 & 0 & 4 \end{bmatrix}$$

and the null space of $\mathbf{A} - \lambda_3\mathbf{I}$ is spanned by

$$\mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

and therefore $\mathbf{v}_3$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ_3. Notice that in this case, the 3×3 matrix $\mathbf{A}$ has three distinct eigenvalues and the eigenvectors

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}$$

correspond to the distinct eigenvalues $\lambda_1, \lambda_2, \lambda_3$, respectively. Therefore, the set $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent (by Theorem 21.6), and therefore β is a basis for $\mathbb{R}^3$. You can verify, for instance, that $\det([\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3]) \neq 0$. □

By Theorem 21.6, the previous example has the following generalization.

Theorem 22.6: Suppose that $\mathbf{A}$ is a $n \times n$ matrix and has n **distinct** eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $\mathbf{v}_i$ be an eigenvector of $\mathbf{A}$ corresponding to λ_i. Then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$.

Hence, if $\mathbf{A}$ has distinct eigenvalues, we are guaranteed the existence of a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$. In forthcoming lectures, we will see that it is very convenient to work with matrices $\mathbf{A}$ that have a set of eigenvectors that form a basis of $\mathbb{R}^n$; this is one of the main motivations for studying eigenvalues and eigenvectors in the first place. However, we will see that not every matrix has a set of eigenvectors that form a basis of $\mathbb{R}^n$. For example, what if $\mathbf{A}$ does not have n distinct eigenvalues? In this case, does there exist a

www.RanaMaths.com

basis for $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}$? In some cases, the answer is yes as the next example demonstrates.

Example 22.7. Find the eigenvalues of $\mathbf{A}$ and a basis for each eigenspace.

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 2 & 2 \\ -2 & 0 & 1 \end{bmatrix}$$

Does $\mathbb{R}^3$ have a basis of eigenvectors of $\mathbf{A}$?

Solution. The characteristic polynomial of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 1)(\lambda - 2)^2$$

and therefore the eigenvalues are $\lambda_1 = 1$ and $\lambda_2 = 2$. Notice that although $p(\lambda)$ is a polynomial of degree $n = 3$, it has **only two distinct roots** and hence $\mathbf{A}$ has only two distinct eigenvalues. The eigenvalue $\lambda_2 = 2$ is said to be **repeated** and $\lambda_1 = 1$ is said to be a **simple** eigenvalue. For $\lambda_1 = 1$ one finds that the eigenspace $\text{Null}(\mathbf{A} - \lambda_1\mathbf{I})$ is spanned by

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$$

and thus $\mathbf{v}_1$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda_1 = 1$. Now consider $\lambda_2 = 2$:

$$\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} 0 & 0 & 0 \\ 4 & 0 & 2 \\ -2 & 0 & -1 \end{bmatrix}$$

Row reducing $\mathbf{A} - 2\mathbf{I}$ one obtains

$$\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} 0 & 0 & 0 \\ 4 & 0 & 2 \\ -2 & 0 & -1 \end{bmatrix} \sim \begin{bmatrix} -2 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Therefore, $\text{rank}(\mathbf{A} - 2\mathbf{I}) = 1$ and thus by the Rank Theorem it follows that $\text{Null}(\mathbf{A} - 2\mathbf{I})$ is a 2-dimensional eigenspace. Performing back substitution, one finds the following basis for the λ_2-eigenspace:

$$\{\mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Therefore, the eigenvectors

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

form a basis for $\mathbb{R}^3$. Hence, for the repeated eigenvalue $\lambda_2 = 2$ we were able to find two linearly independent eigenvectors. $\square$

Before moving further with more examples, we need to introduce some notation regarding the factorization of the characteristic polynomial. In the previous Example 22.7, the characteristic polynomial was factored as $p(\lambda) = (\lambda - 1)(\lambda - 2)^2$ and we found a basis for $\mathbb{R}^3$ of eigenvectors **despite** the presence of a repeated eigenvalue. In general, if $p(\lambda)$ is an nth degree polynomial that can be completely factored into linear terms, then $p(\lambda)$ can be written in the form

$$p(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_p)^{k_p}$$

where $k_1, k_2, \ldots, k_p$ are positive integers and the roots of $p(\lambda)$ are then $\lambda_1, \lambda_2, \ldots, \lambda_k$. Because $p(\lambda)$ is of degree n, we must have that $k_1 + k_2 + \cdots + k_p = n$. Motivated by this, we introduce the following definition.

Definition 22.8: Suppose that $\mathbf{A} \in M_{n\times n}$ has characteristic polynomial $p(\lambda)$ that can be factored as

$$p(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_p)^{k_p}.$$

The exponent k_i is called the **algebraic multiplicity** of the eigenvalue λ_i. The dimension $\text{Null}(\mathbf{A} - \lambda_i \mathbf{I})$ of the eigenspace associated to λ_i is called the **geometric multiplicity** of λ_i.

For simplicity and whenever it is convenient, we will denote the geometric multiplicity of the eigenvalue λ_i as

$$g_i = \dim(\text{Null}(\mathbf{A} - \lambda_i \mathbf{I})).$$

Example 22.9. A 6×6 matrix $\mathbf{A}$ has characteristic polynomial

$$p(\lambda) = \lambda^6 - 4\lambda^5 - 12\lambda^4.$$

Find the eigenvalues of $\mathbf{A}$ and their algebraic multiplicities.

Solution. Factoring $p(\lambda)$ we obtain

$$p(\lambda) = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2)$$

Therefore, the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 0$, $\lambda_2 = 6$, and $\lambda_3 = -2$. Their algebraic multiplicities are $k_1 = 4$, $k_2 = 1$, and $k_3 = 1$, respectively. The eigenvalue $\lambda_1 = 0$ is repeated, while $\lambda_2 = 6$ and $\lambda_3 = -2$ are simple eigenvalues. □

In Example 22.7, we had $p(\lambda) = (\lambda - 1)(\lambda - 2)^2$ and thus $\lambda_1 = 1$ has algebraic multiplicity $k_1 = 1$ and $\lambda_2 = 2$ has algebraic multiplicity $k_2 = 2$. For $\lambda_1 = 1$, we found one linearly independent eigenvector, and therefore λ_1 has geometric multiplicity $g_1 = 1$. For $\lambda_1 = 2$, we found two linearly independent eigenvectors, and therefore λ_2 has geometric multiplicity $g_2 = 2$. However, as we will see in the next example, the geometric multiplicity g_i is in general less than the algebraic multiplicity k_i:

$$g_i \leq k_i$$

www.RanaMaths.com

Example 22.10. Find the eigenvalues of $\mathbf{A}$ and a basis for each eigenspace:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

For each eigenvalue of $\mathbf{A}$, find its algebraic and geometric multiplicity. Does $\mathbb{R}^3$ have a basis of eigenvectors of $\mathbf{A}$?

Solution. One computes

$$p(\lambda) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2$$

and therefore the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 1$ and $\lambda_2 = -2$. The algebraic multiplicity of λ_1 is $k_1 = 1$ and that of λ_2 is $k_2 = 2$. For $\lambda_1 = 1$ we compute

$$\mathbf{A} - \mathbf{I} = \begin{bmatrix} 1 & 4 & 3 \\ -4 & -7 & -3 \\ 3 & 3 & 0 \end{bmatrix}$$

and then one finds that

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

is a basis for the λ_1-eigenspace. Therefore, the geometric multiplicity of λ_1 is $g_1 =$. For $\lambda_2 = -2$ we compute

$$\mathbf{A} - \lambda_2\mathbf{I} = \begin{bmatrix} 4 & 4 & 3 \\ -4 & -4 & -3 \\ 3 & 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 4 & 4 & 3 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore, since $\text{rank}(\mathbf{A} - \lambda_2\mathbf{I}) = 2$, the geometric multiplicity of $\lambda_2 = -2$ is $g_2 = 1$, which is less than the algebraic multiplicity $k_2 = 2$. An eigenvector corresponding to $\lambda_2 = -2$ is

$$\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

Therefore, for the repeated eigenvalue $\lambda_2 = -2$, we are able to find only one linearly independent eigenvector. Therefore, it is not possible to construct a basis for $\mathbb{R}^3$ consisting of eigenvectors of $\mathbf{A}$. □

Hence, in the previous example, there does not exist a basis of $\mathbb{R}^3$ of eigenvectors of $\mathbf{A}$ because for one of the eigenvalues (namely λ_2) the geometric multiplicity was less than the algebraic multiplicity:

$$g_2 < d_2.$$

In the next lecture, we will elaborate on this situation further.

Example 22.11. Find the algebraic and geometric multiplicities of each eigenvalue of the matrix

$$\mathbf{A} = \begin{bmatrix} -7 & 1 & 0 \\ 0 & -7 & 1 \\ 0 & 0 & -7 \end{bmatrix}.$$

www.RanaMaths.com

22.2 Eigenvalues and Similarity Transformations

To end this lecture, we will define a notion of similarity between matrices that plays an important role in linear algebra and that will be used in the next lecture when we discuss diagonalization of matrices. In mathematics, there are many cases where one is interested in classifying objects into categories or classes. Classifying mathematical objects into classes/categories is similar to how some physical objects are classified. For example, all fruits are classified into categories: apples, pears, bananas, oranges, avocados, etc. Given a piece of fruit A, how do you decide what category it is in? What are the properties that uniquely classify the piece of fruit A? In linear algebra, there are many objects of interest. We have spent a lot of time working with matrices and we have now reached a point in our study where we would like to begin classifying matrices. How should we decide if matrices $\mathbf{A}$ and $\mathbf{B}$ are of the same type or, in other words, are similar? Below is how we will decide.

Definition 22.12: Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ matrices. We will say that **A is similar to B** if there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}.$$

If $\mathbf{A}$ is similar to $\mathbf{B}$ then $\mathbf{B}$ is similar to $\mathbf{A}$ because from the equation $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$ we can multiply on the left by $\mathbf{P}^{-1}$ and on the right by $\mathbf{P}$ to obtain that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{B}.$$

Hence, with $\mathbf{Q} = \mathbf{P}^{-1}$, we have that $\mathbf{B} = \mathbf{Q}\mathbf{A}\mathbf{Q}^{-1}$ and thus $\mathbf{B}$ is similar to $\mathbf{A}$. Hence, if $\mathbf{A}$ is similar to $\mathbf{B}$ then $\mathbf{B}$ is similar to $\mathbf{A}$ and therefore we simply say that $\mathbf{A}$ and $\mathbf{B}$ are **similar**. Matrices that are **similar** are clearly not necessarily equal. However, there is a reason why the word **similar** is used. Here are a few reasons why.

Theorem 22.13: If $\mathbf{A}$ and $\mathbf{B}$ are similar matrices then the following are true:
(a) $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B})$
(b) $\det(\mathbf{A}) = \det(\mathbf{B})$
(c) $\mathbf{A}$ and $\mathbf{B}$ have the same eigenvalues

Proof. We will prove part (c). If $\mathbf{A}$ and $\mathbf{B}$ are similar then $\mathbf{A} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$ for some matrix $\mathbf{P}$. Then

$$\begin{aligned}
\det(\mathbf{A} - \lambda\mathbf{I}) &= \det(\mathbf{A} - \lambda\mathbf{P}\mathbf{P}^{-1}) \\
&= \det(\mathbf{P}\mathbf{B}\mathbf{P}^{-1} - \lambda\mathbf{P}\mathbf{P}^{-1}) \\
&= \det(\mathbf{P}(\mathbf{B} - \lambda\mathbf{I})\mathbf{P}^{-1}) \\
&= \det(\mathbf{P})\det(\mathbf{B} - \lambda\mathbf{I})\det(\mathbf{P}^{-1}) \\
&= \det(\mathbf{B} - \lambda\mathbf{I})
\end{aligned}$$

Thus, **A** and **B** have the same characteristic polynomial, and hence the same eigenvalues. □

In the next lecture, we will see that if $\mathbb{R}^n$ has a basis of eigenvectors of **A** then **A** is similar to a diagonal matrix.

After this lecture you should know the following:

- what the characteristic polynomial is and how to compute it
- how to compute the eigenvalues of a matrix
- that when a matrix **A** has distinct eigenvalues, we are guaranteed a basis of $\mathbb{R}^n$ consisting of the eigenvectors of **A**
- that when a matrix **A** has repeated eigenvalues, it is still possible that there exists a basis of $\mathbb{R}^n$ consisting of the eigenvectors of **A**
- what is the algebraic multiplicity and geometric multiplicity of an eigenvalue
- that eigenvalues of a matrix do not change under similarity transformations

www.RanaMaths.com

Lecture 23

Diagonalization

23.1 Eigenvalues of Triangular Matrices

Before discussing diagonalization, we first consider the eigenvalues of triangular matrices.

Theorem 23.1: Let $\mathbf{A}$ be a triangular matrix (either upper or lower). Then the eigenvalues of $\mathbf{A}$ are its diagonal entries.

Proof. We will prove the theorem for the case $n = 3$ and $\mathbf{A}$ is upper triangular; the general case is similar. Suppose then that $\mathbf{A}$ is a 3×3 upper triangular matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$$

Then

$$\mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix}.$$

and thus the characteristic polynomial of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda)$$

and the roots of $p(\lambda)$ are

$$\lambda_1 = a_{11},\ \lambda_2 = a_{22},\ \lambda_3 = a_{33}.$$

In other words, the eigenvalues of $\mathbf{A}$ are simply the diagonal entries of $\mathbf{A}$. □

Example 23.2. Consider the following matrix

$$\mathbf{A} = \begin{bmatrix} 6 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 7 & 0 & 0 \\ -1 & 0 & 0 & -4 & 0 \\ 8 & -2 & 3 & 0 & 7 \end{bmatrix}.$$

www.RanaMaths.com

(a) Find the characteristic polynomial and the eigenvalues of **A**.
(b) Find the geometric and algebraic multiplicity of each eigenvalue of **A**.

We now introduce a very special type of a triangular matrix, namely, a diagonal matrix.

Definition 23.3: A matrix **D** whose off-diagonal entries are all zero is called a diagonal matrix.

For example, here is 3×3 diagonal matrix

$$\mathbf{D} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & -8 \end{bmatrix}.$$

and here is a 5×5 diagonal matrix

$$\mathbf{D} = \begin{bmatrix} 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{7}{2} & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{11} \end{bmatrix}.$$

A diagonal matrix is clearly also a triangular matrix and therefore the eigenvalues of a diagonal matrix **D** are simply the diagonal entries of **D**. Moreover, the powers of a diagonal matrix are easy to compute. For example, if $\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ then

$$\mathbf{D}^2 = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix}$$

and similarly for any integer $k = 1, 2, 3, \ldots$, we have that

$$\mathbf{D}^k = \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}.$$

23.2 Diagonalization

Recall that two matrices **A** and **B** are said to be **similar** if there exists an invertible matrix **P** such that

$$\mathbf{A} = \mathbf{PBP}^{-1}.$$

A very simple type of matrix is a diagonal matrix since many computations with diagonal matrices are trivial. The problem of diagonalization is thus concerned with answering the question of whether a given matrix is similar to a diagonal matrix. Below is the formal definition.

www.RanaMaths.com

Definition 23.4: A matrix $\mathbf{A}$ is called **diagonalizable** if it is similar to a diagonal matrix $\mathbf{D}$. In other words, if there exists an invertible $\mathbf{P}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

How do we determine when a given matrix $\mathbf{A}$ is diagonalizable? Let us first determine what conditions need to be met for a matrix $\mathbf{A}$ to be diagonalizable. Suppose then that $\mathbf{A}$ is diagonalizable. Then by Definition 23.4, there exists an invertible matrix $\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ and a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

such that $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. Multiplying on the right both sides of the equation $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ by the matrix $\mathbf{P}$ we obtain that

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}.$$

Now

$$\mathbf{A}\mathbf{P} = \begin{bmatrix} \mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \cdots & \mathbf{A}\mathbf{v}_n \end{bmatrix}$$

while on the other hand

$$\mathbf{P}\mathbf{D} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}.$$

Therefore, since it holds that $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$ then

$$\begin{bmatrix} \mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \cdots & \mathbf{A}\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}.$$

or if we compare columns we must have that

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i.$$

Thus, the columns $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of $\mathbf{P}$ are eigenvectors of $\mathbf{A}$ and form a basis for $\mathbb{R}^n$ because $\mathbf{P}$ is invertible. In conclusion, if $\mathbf{A}$ is diagonalizable then $\mathbb{R}^n$ has a basis consisting of eigenvectors of $\mathbf{A}$.

Suppose instead that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$ associated to $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, respectively, and set

$$\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

Then $\mathbf{P}$ is invertible because $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are linearly independent. Let

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}.$$

Now, since $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ we have that

$$\begin{aligned}\mathbf{AP} &= \mathbf{A}\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \cdots & \mathbf{A}\mathbf{v}_n\end{bmatrix} \\ &= \begin{bmatrix}\lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\end{bmatrix}.\end{aligned}$$

Therefore, $\mathbf{AP} = \begin{bmatrix}\lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\end{bmatrix}$. On the other hand,

$$\begin{aligned}\mathbf{PD} &= \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \ldots & 0\\ 0 & \lambda_2 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \lambda_n\end{bmatrix} \\ &= \begin{bmatrix}\lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\end{bmatrix}.\end{aligned}$$

Therefore, $\mathbf{AP} = \mathbf{PD}$, and since $\mathbf{P}$ is invertible we have that

$$\mathbf{A} = \mathbf{PDP}^{-1}.$$

Thus, if $\mathbb{R}^n$ has a basis of consisting of eigenvectors of $\mathbf{A}$ then $\mathbf{A}$ is diagonalizable. We have therefore proved the following theorem.

Theorem 23.5: A matrix $\mathbf{A}$ is diagonalizable if and only if there is a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.

The punchline with Theorem 23.5 is that the problem of diagonalization of a matrix $\mathbf{A}$ is equivalent to finding a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$. We will see in some of the examples below that it is not always possible to diagonalize a matrix.

23.3 Conditions for Diagonalization

We first consider the simplest case when we conclude that a given matrix is diagonalizable, namely, the case when all eigenvalues are distinct.

Theorem 23.6: Suppose that $\mathbf{A} \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then $\mathbf{A}$ is diagonalizable.

Proof. Each eigenvalue λ_i produces an eigenvector $\mathbf{v}_i$. The set of eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are linearly independent because they correspond to distinct eigenvalues (Theorem 21.6). Therefore, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$ and then by Theorem 23.5 we conclude that $\mathbf{A}$ is diagonalizable. □

What if $\mathbf{A}$ does not have distinct eigenvalues? Can $\mathbf{A}$ still be diagonalizable? The following theorem completely answers this question.

www.RanaMaths.com

Theorem 23.7: A matrix $\mathbf{A}$ is diagonalizable if and only if the algebraic and geometric multiplicities of each eigenvalue are equal.

Proof. Let $\mathbf{A}$ be a $n \times n$ matrix and let $\lambda_1, \lambda_2, \ldots, \lambda_p$ denote the distinct eigenvalues of $\mathbf{A}$. Suppose that $k_1, k_2, \ldots, k_p$ are the algebraic multiplicities and $g_1, g_2, \ldots, g_p$ are the geometric multiplicities of the eigenvalues, respectively. Suppose that the algebraic and geometric multiplicities of each eigenvalue are equal, that is, suppose that $g_i = k_i$ for each $i = 1, 2 \ldots, p$. Since $k_1 + k_2 + \cdots + k_p = n$, then because $g_i = k_i$ we must also have that $g_1 + g_2 + \cdots + g_p = n$. Therefore, there exists n linearly eigenvectors of $\mathbf{A}$ and consequently $\mathbf{A}$ is diagonalizable. On the other hand, suppose that $\mathbf{A}$ is diagonalizable. Since the geometric multiplicity is at most the algebraic multiplicity, the only way that $g_1 + g_2 + \cdots + g_p = n$ is if $g_i = k_i$, i.e., that the geometric and algebraic multiplicities are equal. □

Example 23.8. Determine if $\mathbf{A}$ is diagonalizable. If yes, find a matrix $\mathbf{P}$ that diagonalizes $\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} -4 & -6 & -7 \\ 3 & 5 & 3 \\ 0 & 0 & 3 \end{bmatrix}$$

Solution. The characteristic polynomial of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = (\lambda - 2)(\lambda - 3)(\lambda + 1)$$

and therefore $\lambda_1 = 2$, $\lambda_2 = 3$, and $\lambda_3 = -1$ are the eigenvalues of $\mathbf{A}$. Since $\mathbf{A}$ has $n = 3$ distinct eigenvalues, then by Theorem 23.6 $\mathbf{A}$ is diagonalizable. Eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ corresponding to $\lambda_1, \lambda_2, \lambda_3$ are found to be

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

Therefore, a matrix that diagonalizes $\mathbf{A}$ is

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

You can verify that

$$\mathbf{P} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \mathbf{P}^{-1} = \mathbf{A}$$

□

The following example demonstrates that it is possible for a matrix to be diagonalizable even though the matrix does not have distinct eigenvalues.

www.RanaMaths.com

Example 23.9. Determine if $\mathbf{A}$ is diagonalizable. If yes, find a matrix $\mathbf{P}$ that diagonalizes $\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 4 & 2 & 2 \\ -2 & 0 & 1 \end{bmatrix}$$

Solution. The characteristic polynomial of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda - 1)(\lambda - 2)^2$$

and therefore $\lambda_1 = 1$, $\lambda_2 = 2$. An eigenvector corresponding to $\lambda_1 = 1$ is

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$$

One finds that $g_2 = \dim(\text{Null}(\mathbf{A} - \lambda_2\mathbf{I})) = 2$, and two linearly independent eigenvectors for λ_2 are

$$\{\mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Therefore, $\mathbf{A}$ is diagonalizable, and a matrix that diagonalizes $\mathbf{A}$ is

$$\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ -2 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix}$$

You can verify that

$$\mathbf{P} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \mathbf{P}^{-1} = \mathbf{A}$$

□

Example 23.10. Determine if $\mathbf{A}$ is diagonalizable. If yes, find a matrix $\mathbf{P}$ that diagonalizes $\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

Solution. The characteristic polynomial of $\mathbf{A}$ is

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2$$

and therefore the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 1$ and $\lambda_2 = -2$. For $\lambda_2 = -2$ one computes

$$\mathbf{A} - \lambda_2\mathbf{I} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

We see that the dimension of the eigenspace of $\lambda_2 = -2$ is $g_2 = 1$, which is less than the algebraic multiplicity $k_2 = 2$. Therefore, from Theorem 23.7 we can conclude that it is not possible to construct a basis of eigenvectors of $\mathbf{A}$, and therefore $\mathbf{A}$ is not diagonalizable. □

www.RanaMaths.com

Example 23.11. Suppose that $\mathbf{A}$ has eigenvector $\mathbf{v}$ with corresponding eigenvalue λ. Show that if $\mathbf{A}$ is invertible then $\mathbf{v}$ is an eigenvector of $\mathbf{A}^{-1}$ with corresponding eigenvalue $\frac{1}{\lambda}$.

Example 23.12. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices such that $\mathbf{AB} = \mathbf{BA}$. Show that if $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ with corresponding eigenvalue then $\mathbf{v}$ is also an eigenvector of $\mathbf{B}$ with corresponding eigenvalue λ.

After this lecture you should know the following:

- Determine if a matrix is diagonalizable or not
- Find the algebraic and geometric multiplicities of an eigenvalue
- Apply the theorems introduced in this lecture

www.RanaMaths.com

Lecture 24

Diagonalization of Symmetric Matrices

24.1 Symmetric Matrices

Recall that a square matrix $\mathbf{A}$ is said to be **symmetric** if $\mathbf{A}^T = \mathbf{A}$. As an example, here is a 3×3 symmetric matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & -3 & 7 \\ -3 & 2 & 8 \\ 7 & 8 & 4 \end{bmatrix}.$$

Symmetric matrices are ubiquitous in mathematics. For example, let $f(x_1, x_2, \ldots, x_n)$ be a function having continuous second order partial derivatives. Then Clairaut's Theorem from multivariable calculus says that

$$\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial f}{\partial x_j \partial x_i}.$$

Therefore, the **Hessian** matrix of f is symmetric:

$$\text{Hess}(f) = \begin{bmatrix} \frac{\partial f}{\partial x_1 \partial x_1} & \frac{\partial f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial f}{\partial x_1 \partial x_n} \\ \frac{\partial f}{\partial x_2 \partial x_1} & \frac{\partial f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_n \partial x_1} & \frac{\partial f}{\partial x_n \partial x_2} & \cdots & \frac{\partial f}{\partial x_n \partial x_n} \end{bmatrix}.$$

The **Second Derivative Test** of multivariable calculus then says that if $P = (a_1, a_2, \ldots, a_n)$ is a critical point of f, that is

$$\frac{\partial f}{\partial x_1}(P) = \frac{\partial f}{\partial x_2}(P) = \cdots = \frac{\partial f}{\partial x_n}(P) = 0$$

then

(i) P is a local minimum point of f if the matrix $\text{Hess}(f)$ has all positive eigenvalues,

(ii) P is a local maximum point of f if the matrix $\text{Hess}(f)$ has all negative eigenvalues, and

www.RanaMaths.com

(iii) P is a saddle point of f if the matrix Hess(f) has negative and positive eigenvalues.

In general, the eigenvalues of a matrix with real entries can be complex numbers. For example, the matrix

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

has characteristic polynomial

$$p(\lambda) = \lambda^2 + 1$$

the roots of which are clearly $\lambda_1 = \sqrt{-1} = i$ and $\lambda_2 = -\sqrt{-1} = -i$. Thus, in general, a matrix whose entries are all real numbers may have complex eigenvalues. However, for symmetric matrices we have the following.

Theorem 24.1: If $\mathbf{A}$ is a symmetric matrix then all of its eigenvalues are real numbers.

The proof is easy but we will omit it.

24.2 Eigenvectors of Symmetric Matrices

We proved earlier that if $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ are eigenvectors of a matrix $\mathbf{A}$ corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ then the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly independent (Theorem 21.6). For symmetric matrices we can say even more as the next theorem states.

Theorem 24.2: Let $\mathbf{A}$ be a symmetric matrix. If $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of $\mathbf{A}$ corresponding to distinct eigenvalues then $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal, that is, $\mathbf{v}_1 \bullet \mathbf{v}_2 = 0$.

Proof. Recall that $\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{v}_2$. Let $\lambda_1 \neq \lambda_2$ be the eigenvalues associated to $\mathbf{v}_1$ and $\mathbf{v}_2$. Then

$$\begin{aligned} \lambda_1 \mathbf{v}_1^T \mathbf{v}_2 &= (\lambda_1 \mathbf{v}_1)^T \mathbf{v}_2 \\ &= (\mathbf{A}\mathbf{v}_1)^T \mathbf{v}_2 \\ &= \mathbf{v}_1^T \mathbf{A}^T \mathbf{v}_2 \\ &= \mathbf{v}_1^T \mathbf{A} \mathbf{v}_2 \\ &= \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) \\ &= \lambda_2 \mathbf{v}_1^T \mathbf{v}_2. \end{aligned}$$

Therefore, $\lambda_1 \mathbf{v}_1^T \mathbf{v}_2 = \lambda_2 \mathbf{v}_1^T \mathbf{v}_2$ which implies that $(\lambda_1 - \lambda_2)\mathbf{v}_1^T \mathbf{v}_2 = 0$. But since $(\lambda_1 - \lambda_2) \neq 0$ then we must have $\mathbf{v}_1^T \mathbf{v}_2 = 0$, that is, $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. $\square$

24.3 Symmetric Matrices are Diagonalizable

As we have seen, the main criteria for diagonalization is that for each eigenvalue the geometric and algebraic multiplicities are equal; not all matrices satisfy this condition and thus not

www.RanaMaths.com

all matrices are diagonalizable. As it turns out, any **symmetric A** is diagonalizable and moreover (and perhaps more importantly) there exists an **orthogonal** eigenvector matrix **P** that diagonalizes **A**. The full statement is below.

Theorem 24.3: If **A** is a symmetric matrix then **A** is diagonalizable. In fact, there is an orthonormal basis of $\mathbb{R}^n$ of eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of **A**. In other words, the matrix $\mathbf{P} = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n]$ is orthogonal, $\mathbf{P}^T\mathbf{P} = \mathbf{I}$, and $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$.

The proof of the theorem is not hard but we will omit it. The punchline of Theorem 24.3 is that, for the case of a symmetric matrix, we will never encounter the situation where the geometric multiplicity is strictly less than the algebraic multiplicity. Moreover, we are guaranteed to find an orthogonal matrix that diagonalizes a given symmetric matrix.

Example 24.4. Find an orthogonal matrix **P** that diagonalizes the symmetric matrix

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

Solution. The characteristic polynomial of **A** is

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \lambda^3 - 4\lambda^2 + 3\lambda = \lambda(\lambda - 1)(\lambda - 3)$$

The eigenvalues of **A** are $\lambda_1 = 0$, $\lambda_2 = 1$ and $\lambda_3 = 3$. Eigenvectors of **A** associated to $\lambda_1, \lambda_2, \lambda_3$ are

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}.$$

As expected by Theorem 24.2, the eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ form an orthogonal set:

$$\mathbf{u}_1^T\mathbf{u}_2 = 0,\ \mathbf{u}_1^T\mathbf{u}_3 = 0,\ \mathbf{u}_2^T\mathbf{u}_3 = 0.$$

To find an orthogonal matrix **P** that diagonalizes **A** we must normalize the eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ to obtain an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. To that end, first compute $\mathbf{u}_1^T\mathbf{u}_1 = 3$, $\mathbf{u}_2^T\mathbf{u}_2 = 2$, and $\mathbf{u}_3^T\mathbf{u}_3 = 6$. Then let $\mathbf{v}_1 = \frac{1}{\sqrt{3}}\mathbf{u}_1$, let $\mathbf{v}_2 = \frac{1}{\sqrt{2}}\mathbf{u}_2$, and let $\mathbf{v}_3 = \frac{1}{\sqrt{6}}\mathbf{u}_3$. Therefore, an orthogonal matrix that diagonalizes **A** is

$$\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}$$

You can easily verify that $\mathbf{P}^T\mathbf{P} = \mathbf{I}$, and that

$$\mathbf{A} = \mathbf{P} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{P}^T$$

□

www.RanaMaths.com

Example 24.5. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ matrices. Show that if $\mathbf{A}$ is symmetric then the matrix $\mathbf{C} = \mathbf{B}\mathbf{A}\mathbf{B}^T$ is also a symmetric matrix.

After this lecture you should know the following:

- a symmetric matrix is diagonalizable with an orthonormal set of eigenvectors

Lecture 25

The PageRank Algortihm

In this lecture, we will see how linear algebra is used in Google's webpage ranking algorithm used in everyday Google searches.

25.1 Search Engine Retrieval Process

Search engines perform a two-stage process to retrieve search results[1]. In Stage 1, traditional text processing is used to find all relevant pages (e.g. keywords in title, body) and produces a *content score*. After **Stage 1**, there is a large amount of relevant pages. For example, the query "*symmetric matrix*" results in about 3,830,000 pages (03/31/15). Or "*homework help*" results in 49,400,000 pages (03/31/15). How should the relevant pages be displayed? In Stage 2, the pages are sorted and displayed based on a pre-computed ranking that is **query-independent**, this is the *popularity score*. The ranking is based on the hyperlinked or networked structure of the web, and the ranking is based on a popularity contest; if many pages link to page P_i then P_i must be an important page and should therefore have a high popularity score.

In January 1998, John Kleinberg from IBM (now a CS professor at Cornell) presented the **HITS** algorithm[2] (e.g., `www.teoma.com`). At Stanford, doctoral students Sergey Brin and Larry Page were busy working on a similar project which they had begun in 1995. Below is the abstract of their paper[3]:

"In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/ ."

[1] A.N. Langville and C.D. Meyer, *Google's PageRank and Beyond*, Princeton University Press, 2006

[2] J. Kleinberg, *Authoritative sources in a hyperlinked environment*, Journal of ACM, 46, 1999, 9th ACM-SIAM Symposium on Discrete Algorithms

[3] S. Brin and L. Page, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks and ISDN Systems, 33:107-117, 1998

www.RanaMaths.com

In both models, the web is defined as a **directed graph**, where the nodes represent webpages and the directed arcs represent hyperlinks, see Figure 25.1.

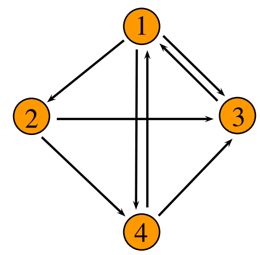


Figure 25.1: A tiny web represented as a directed graph.

25.2 A Description of the PageRank Algorithm

In the PageRank algorithm, each inlink is viewed as a recommendation (or vote). In general, pages with many inlinks are more important than pages with few inlinks. However, the *quality* of the inlink (vote) is important. The vote of each page should be divided by the total number of recommendations made by the page. The **PageRank** of page i, denoted x_i, is the sum of all the weighted PageRanks of all the pages pointing to i:

$$x_i = \sum_{j \to i} \frac{x_j}{|N_j|}$$

where

(1) N_j is the number of outlinks from page j

(2) $j \to i$ means page j links to page i

Example 25.1. Find the PageRank of each page for the network in Figure 25.1.

From the previous example, we see that the PageRank of each page can be found by solving an eigenvalue/eigenvector problem. However, when dealing with large networks such as the internet, the size of the problem is in the billions (8.1 billion in 2006) and directly solving the equations is not possible. Instead, an iterative method called the **power method** is used. One starts with an initial guess, say $\mathbf{x}_0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Then one updates the guess by computing

$$\mathbf{x}_1 = \mathbf{H}\mathbf{x}_0.$$

In other words, we have a discrete dynamical system

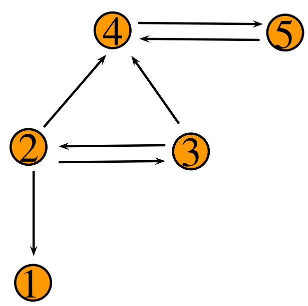
$$\mathbf{x}_{k+1} = \mathbf{H}\mathbf{x}_k.$$

A natural question is under what conditions will the the limiting value of the sequence

$$\lim_{k\to\infty} \mathbf{x}_k = \lim_{k\to\infty} (\mathbf{H}^k \mathbf{x}_0) = \mathbf{q}$$

www.RanaMaths.com

converge to an equilibrium of $\mathbf{H}$? Also, if $\lim_{k\to\infty} \mathbf{x}_k$ exists, will it be a positive vector? And lastly, can $\mathbf{x}_0 \neq \mathbf{0}$ be chosen arbitrarily? To see what situations may occur, consider the network displayed in Figure 25.2. Starting with $\mathbf{x}_0 = (\frac{1}{5}, \ldots, \frac{1}{5})$ we obtain that for $k \geq 39$, the vectors $\mathbf{x}_k = \mathbf{H}^k \mathbf{x}_0$ cycle between $(0, 0, 0, 0.28, 0.40)$ and $(0, 0, 0, 0.40, 0.28)$. Therefore, the sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$ does not converge. The reason for this is that nodes 4 and 5 form a **cycle**.



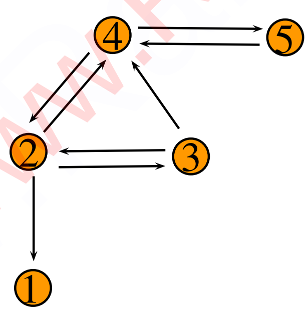
$$\mathbf{H} = \begin{bmatrix} 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Figure 25.2: Cycles present in the network

Now consider the network displayed in Figure 25.3. If we remove the cycle we are still left with a **dangling node**, namely node 1 (e.g. pdf file, image file). Starting with $\mathbf{x}_0 = (\frac{1}{5}, \ldots, \frac{1}{5})$ results in

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{0}.$$

Therefore, in this case the sequence $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$ converges to a non-positive vector, which for the purposes of ranking pages would be an undesirable situation.



$$\mathbf{H} = \begin{bmatrix} 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

Figure 25.3: Dangling node present in the network

To avoid the presence of dangling nodes and cycles, Brin and Page used the notion of a **random surfer** to adjust $\mathbf{H}$. To deal with a dangling node, Brin and Page replaced the associated zero-column with the vector $\frac{1}{n}\mathbf{1} = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$. The justification for this adjustment is that if a random surfer reaches a dangling node, the surfer will "teleport" to any page in the web with equal probability. The new updated hyperlink matrix $\mathbf{H}^*$ may still not have the desired properties. To deal with cycles, a surfer may abandon the hyperlink structure of the web by ocassionally moving to a random page by typing its address in the

www.RanaMaths.com

browser. With these adjustments, a random surfer now spends only a proportion of his time using the hyperlink structure of the web to visit pages. Hence, let $0 < \alpha < 1$ be the proportion of time the random surfer uses the hyperlink structure. Then the transition matrix is

$$\mathbf{G} = \alpha \mathbf{H}^* + (1-\alpha)\tfrac{1}{n}\mathbf{J}.$$

The matrix $\mathbf{G}$ goes by the name of the Google matrix, and it is reported that Google uses $\alpha = 0.85$ (here $\mathbf{J}$ is the all ones matrix). The Google matrix $\mathbf{G}$ is now a **primitive** and **stochastic** matrix. **Stochastic** means that all its columns are probability vectors, i.e., non-negative vectors whose components sum to 1. **Primitive** means that there exists $k \geq 1$ such that $\mathbf{G}^k$ has all positive entries ($k = 1$ in our case). With these definitions, we now have the following theorem.

Theorem 25.2: If $\mathbf{G}$ is a primitive stochastic matrix then:

(i) There is a stochastic $\mathbf{G}^*$ such that $\lim_{k\to\infty} \mathbf{G}^k = \mathbf{G}^*$.

(ii) $\mathbf{G}^* = \begin{bmatrix} \mathbf{q} & \mathbf{q} & \cdots & \mathbf{q} \end{bmatrix}$ where $\mathbf{q}$ is a probability vector.

(iii) For any probability vector $\mathbf{q}_0$ we have $\lim_{k\to\infty} \mathbf{G}^k \mathbf{q}_0 = \mathbf{q}$.

(iv) The vector $\mathbf{q}$ is the unique probability vector which is an eigenvector of $\mathbf{G}$ with eigenvalue $\lambda_1 = 1$.

(v) All other eigenvalues $\lambda_2, \ldots, \lambda_n$ have $|\lambda_j| < 1$.

Proof. We will prove a special case[4]. Assume for simplicity that $\mathbf{G}$ is positive (this is the case of the Google Matrix). If $\mathbf{x} = \mathbf{G}\mathbf{x}$, and $\mathbf{x}$ has mixed signs, then

$$|x_i| = \left| \sum_{j=1}^n \mathbf{G}_{ij} x_j \right| < \sum_{j=1}^n \mathbf{G}_{ij} |x_j|.$$

Then

$$\sum_{i=1}^n |x_i| < \sum_{i=1}^n \sum_{j=1}^n \mathbf{G}_{ij} |x_j| = \sum_{j=1} |x_j|$$

which is a contradiction. Therefore, all the eigenvectors in the $\lambda_1 = 1$ eigenspace are either negative or positive. One then shows that the eigenspace corresponding to $\lambda_1 = 1$ is 1-dimensional. This proves that there is a unique probability vector $\mathbf{q}$ such that

$$\mathbf{q} = \mathbf{G}\mathbf{q}.$$

[4] K. Bryan, T. Leise, *The \$25,000,000,000 Eigenvector: The Linear Algebra Behind Google*, SIAM Review, 48(3), 569-581

www.RanaMaths.com

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{G}$. We know that $\lambda_1 = 1$ is a dominant eigenvalue:

$$|\lambda_1| > |\lambda_j|, \quad j = 2, 3, \ldots, n.$$

Let $\mathbf{q}_0$ be a probability vector and let $\mathbf{q}$ be as above, and let $\mathbf{v}_2, \ldots, \mathbf{v}_n$ be the remaining eigenvectors of $\mathbf{G}$. Then $\mathbf{q}_0 = \mathbf{q} + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ and therefore

$$\begin{aligned}
\mathbf{G}^k\mathbf{q}_0 &= \mathbf{G}^k(\mathbf{q} + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) \\
&= \mathbf{G}^k\mathbf{q} + c_2\mathbf{G}^k\mathbf{v}_2 + \cdots + c_n\mathbf{G}^k\mathbf{v}_n \\
&= \mathbf{q} + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_n\lambda_n^k\mathbf{v}_n.
\end{aligned}$$

From this we see that

$$\lim_{k\to\infty} \mathbf{G}^k\mathbf{q}_0 = \mathbf{q}.$$

□

25.3 Computation of the PageRank Vector

The Google matrix $\mathbf{G}$ is completely dense, which is computationally undesirable. Fortunately,

$$\begin{aligned}
\mathbf{G} &= \alpha\mathbf{H}^* + (1-\alpha)\tfrac{1}{n}\mathbf{e}\mathbf{e}^T \\
&= \alpha(\mathbf{H} + \tfrac{1}{n}\mathbf{1}\mathbf{1}^T) + (1-\alpha)\tfrac{1}{n}\mathbf{1}\mathbf{1}^T \\
&= \alpha\mathbf{H} + (\alpha\mathbf{a} + (1-\alpha)\mathbf{1})\tfrac{1}{n}\mathbf{1}^T
\end{aligned}$$

and $\mathbf{H}$ is very sparse and requires minimal storage. A vector-matrix multiplication generally requires $O(n^2)$ computation ($n \approx 8,000,000,000$ in 2006). Estimates show that the average webpage has about 10 outlinks, so $\mathbf{H}$ has about $10n$ non-zero entries. This means that multiplication with $\mathbf{H}$ reduces to $O(n)$ computation. Aside from being very simple, the power method is a **matrix-free** method, i.e., no manipulation of the matrix $\mathbf{H}$ is done. Brin and Page, and others, have confirmed that only 50-100 iterations are needed for a satisfactory approximation of the PageRank vector $\mathbf{q}$ for the web.

After this lecture you should know the following:

- Setup a Google matrix and compute PageRank vector

Lecture 26

Discrete Dynamical Systems

26.1 Discrete Dynamical Systems

Many interesting problems in engineering, science, and mathematics can be studied within the framework of discrete dynamical systems. Dynamical systems are used to model systems that change over time. The state of the system (economic, ecologic, engineering, etc.) is measured at discrete time intervals producing a sequence of vectors $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$. The relationship between the vector $\mathbf{x}_k$ and the next vector $\mathbf{x}_{k+1}$ is what constitutes a **model**.

Definition 26.1: A **linear discrete dynamical system** on $\mathbb{R}^n$ is an infinite sequence $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots\}$ of vectors in $\mathbb{R}^n$ and a matrix $\mathbf{A}$ such that

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k.$$

The vectors $\mathbf{x}_k$ are called the **state** of the dynamical system and $\mathbf{x}_0$ is the initial condition of the system. Once the initial condition $\mathbf{x}_0$ is fixed, the remaining state vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots$, can be found by iterating the equation $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$.

26.2 Population Model

Consider the dynamic system consisting of the population movement between a city and its suburbs. Let $\mathbf{x} \in \mathbb{R}^2$ be the state population vector whose first component is the population of the city and the second component is the population of the suburbs:

$$\mathbf{x} = \begin{bmatrix} c \\ s \end{bmatrix}.$$

For simplicity, we assume that $c + s = 1$, i.e., c and s are population percentages of the total population. Suppose that in the year 1900, the city population was c_0 and the suburban population was s_0. Suppose it is known that after each year 5% of the city's population

www.RanaMaths.com

moves to the suburbs and that 3% of the suburban population moves to the city. Hence, the population in the city in year 1901 is

$$c_1 = 0.95c_0 + 0.03s_0,$$

while the population in the suburbs in year 1901 is

$$s_1 = 0.05c_0 + 0.97s_0.$$

The equations

$$\begin{aligned} c_1 &= 0.95c_0 + 0.03s_0 \\ s_1 &= 0.05c_0 + 0.97s_0 \end{aligned}$$

can be written in matrix form as

$$\begin{bmatrix} c_1 \\ s_1 \end{bmatrix} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} c_0 \\ s_0 \end{bmatrix}.$$

Performing the same analysis for the next year, the population in 1902 is

$$\begin{bmatrix} c_2 \\ s_2 \end{bmatrix} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} c_1 \\ s_1 \end{bmatrix}.$$

Hence, the population movement is a linear dynamical system with matrix and state vector

$$\mathbf{A} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}, \quad \mathbf{x}_k = \begin{bmatrix} c_k \\ s_k \end{bmatrix}.$$

Suppose that the initial population state vector is

$$\mathbf{x}_0 = \begin{bmatrix} 0.70 \\ 0.30 \end{bmatrix}.$$

Then,

$$\mathbf{x}_1 = \mathbf{A}\mathbf{x}_0 = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} 0.70 \\ 0.30 \end{bmatrix} = \begin{bmatrix} 0.674 \\ 0.326 \end{bmatrix}.$$

Then,

$$\mathbf{x}_2 = \mathbf{A}\mathbf{x}_1 = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix} \begin{bmatrix} 0.674 \\ 0.326 \end{bmatrix} = \begin{bmatrix} 0.650 \\ 0.349 \end{bmatrix}.$$

In a similar fashion, one can compute that up to 3 decimal places:

$$\mathbf{x}_{500} = \begin{bmatrix} 0.375 \\ 0.625 \end{bmatrix}, \quad \mathbf{x}_{1000} = \begin{bmatrix} 0.375 \\ 0.625 \end{bmatrix}.$$

It seems as though the population distribution converges to a steady state or equilibrium. We predict that in the year 2400, 38% of the total population will live in the city and 62% in the suburbs.

Our computations in the population model indicate that the population distribution is reaching a sort of steady state or equilibrium, which we now define.

www.RanaMaths.com

Definition 26.2: Let $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ be a discrete dynamical system. An **equilibrium** state for $\mathbf{A}$ is a vector $\mathbf{q}$ such that $\mathbf{A}\mathbf{q} = \mathbf{q}$.

Hence, if $\mathbf{q}$ is an equilibrium for $\mathbf{A}$ and the initial condition is $\mathbf{x}_0 = \mathbf{q}$ then $\mathbf{x}_1 = \mathbf{A}\mathbf{x}_0 = \mathbf{x}_0$, and $\mathbf{x}_2 = \mathbf{A}\mathbf{x}_1 = \mathbf{x}_0$, and iteratively we have that $\mathbf{x}_k = \mathbf{x}_0 = \mathbf{q}$ for all k. Thus, if the system starts at the equilibrium $\mathbf{q}$ then it remains at $\mathbf{q}$ for all time.

How do we find equilibrium states? If $\mathbf{q}$ is an equilibrium for $\mathbf{A}$ then from $\mathbf{A}\mathbf{q} = \mathbf{q}$ we have that

$$\mathbf{A}\mathbf{q} - \mathbf{q} = \mathbf{0}$$

and therefore

$$(\mathbf{A} - \mathbf{I})\mathbf{q} = \mathbf{0}.$$

Therefore, $\mathbf{q}$ is an equilibrium for $\mathbf{A}$ if and only if $\mathbf{q}$ is in the nullspace of the matrix $\mathbf{A} - \mathbf{I}$:

$$\mathbf{q} \in \text{Null}(\mathbf{A} - \mathbf{I}).$$

Example 26.3. Find the equilibrium states of the matrix from the population model

$$\mathbf{A} = \begin{bmatrix} 0.95 & 0.03 \\ 0.05 & 0.97 \end{bmatrix}.$$

Does the initial condition of the population $\mathbf{x}_0$ change the long term behavior of the discrete dynamical system? We will know the answer once we perform an eigenvalue analysis on $\mathbf{A}$ (Lecture 22). As a preview, we will use the fact that

$$\mathbf{x}_k = \mathbf{A}^k\mathbf{x}_0$$

and then write $\mathbf{x}_0$ in an appropriate basis that reveals how $\mathbf{A}$ acts on $\mathbf{x}_0$. To see how the last equation was obtained, notice that

$$\mathbf{x}_1 = \mathbf{A}\mathbf{x}_0$$

and therefore

$$\mathbf{x}_2 = \mathbf{A}\mathbf{x}_1 = \mathbf{A}(\mathbf{A}\mathbf{x}_0) = \mathbf{A}^2\mathbf{x}_0$$

and therefore

$$\mathbf{x}_3 = \mathbf{A}\mathbf{x}_2 = \mathbf{A}(\mathbf{A}^2\mathbf{x}_0) = \mathbf{A}^3\mathbf{x}_0$$

etc.

26.3 Stability of Discrete Dynamical Systems

We first formally define the notion of stability of a discrete dynamical system.

www.RanaMaths.com

Definition 26.4: Consider the discrete dynamical system $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ where $\mathbf{A} \in \mathbb{R}^{n\times n}$. The origin $\mathbf{0} \in \mathbb{R}^n$ is said to be **asymptotically stable** if for any initial condition $\mathbf{x}_0 \in \mathbb{R}^n$ of the dynamical system we have

$$\lim_{k\to\infty} \mathbf{x}_k = \lim_{k\to\infty} \mathbf{A}^k \mathbf{x}_0 = \mathbf{0}.$$

The following theorem characterizes when a discrete linear dynamical system is asymptotically stable.

Theorem 26.5: Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\mathbf{A}$. If $|\lambda_j| < 1$ for all $j = 1, 2, \ldots, n$ then the origin $\mathbf{0}$ is asymptotically stable for $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$.

Solution. For simplicity, we suppose that $\mathbf{A}$ is diagonalizable. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of eigenvectors of $\mathbf{A}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ respectively. Then, for any vector $\mathbf{x}_0 \in \mathbb{R}^n$, there exists constants $c_1, \ldots, c_n$ such that

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

Now, for any integer $k \geq 1$ we have that.

$$\mathbf{A}^k \mathbf{v}_i = \lambda_i^k \mathbf{v}_i$$

Then

$$\begin{aligned}
\mathbf{x}_k &= \mathbf{A}^k \mathbf{x}_0 \\
&= \mathbf{A}^k (c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\
&= c_1\mathbf{A}^k\mathbf{v}_1 + \cdots + c_n\mathbf{A}^k\mathbf{v}_n \\
&= c_1\lambda_1^k\mathbf{v}_1 + \cdots + c_n\lambda_n^k\mathbf{v}_n.
\end{aligned}$$

Since $|\lambda_i| < 1$ we have that $\lim_{k\to\infty} \lambda_i^k = 0$. Therefore,

$$\begin{aligned}
\lim_{k\to\infty} \mathbf{x}_k &= \lim_{k\to\infty} (c_1\lambda_1^k\mathbf{v}_1 + \cdots + c_n\lambda_n^k\mathbf{v}_n) \\
&= c_1 \left(\lim_{k\to\infty} \lambda_1^k\right) \mathbf{v}_1 + \cdots + c_n \left(\lim_{k\to\infty} \lambda_n^k\right) \mathbf{v}_n \\
&= 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n \\
&= \mathbf{0}.
\end{aligned}$$

This completes the proof. □

www.RanaMaths.com

As an example of an asymptotically stable dynamical system, consider the 2D system

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1.1 & -0.4 \\ 0.15 & 0.6 \end{bmatrix} \mathbf{x}.$$

The eigenvalues of $\mathbf{A} = \begin{bmatrix} 1.1 & -0.4 \\ 0.15 & 0.6 \end{bmatrix}$ are $\lambda_1 = 0.8$ and $\lambda_2 = 0.9$. Hence, by Theorem 26.5, for any initial condition $\mathbf{x}_0$, the sequence $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \}$ converges to the origin in $\mathbb{R}^2$. In Figure 26.1, we plot four different state sequences $\{\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots, \}$ corresponding to the four distinct initial conditions $\mathbf{x}_0 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} 3 \\ -7 \end{bmatrix}$, $\mathbf{x}_0 = \begin{bmatrix} -3 \\ 7 \end{bmatrix}$, and $\mathbf{x}_0 = \begin{bmatrix} -3 \\ -7 \end{bmatrix}$. As expected, all trajectories converge to the origin.

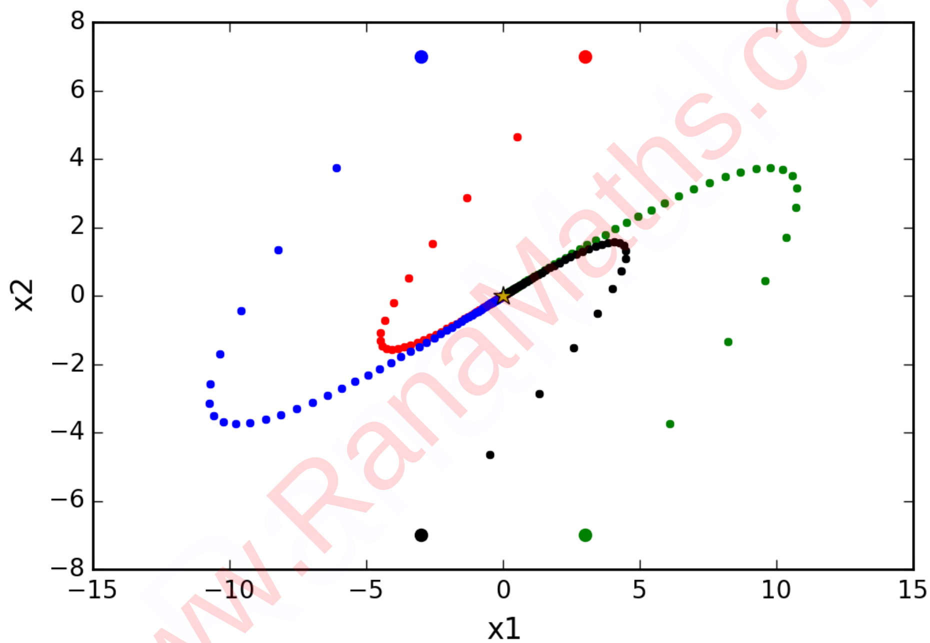


Figure 26.1: A 2D asymptotically stable linear system

After this lecture you should know the following:

- what a dynamical system is
- and how to find its equilibrium states
- how to determine if a discrete dynamical system has the origin as an asymptotically stable equilibrium