Topic 1

Vectors in $\mathbb{R}^n$ (often $\mathbb{R}^2$ and $\mathbb{R}^3$ for simplicity)

Definition 1: Vector

We use the word **vector** to mean that an object has both **magnitude** and **direction**.

Examples 1

Your height is just a number, it has no direction, and so it is not a vector.

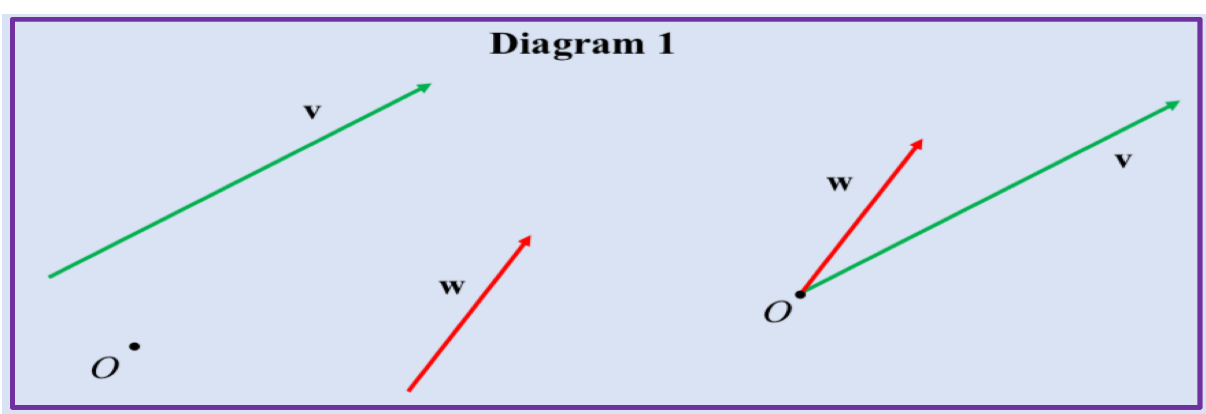
The velocity of a boat on a river is a vector. We are concerned with the speed of the boat, which is the magnitude of the velocity. We would also like to know about the direction that the boat is moving in. The velocity (vector) of the boat contains both of these pieces of information.

Notation 1
In the notes I will use bold text to indicate a vector: $\mathbf{v}$.
When I am writing, as writing in bold cannot be done, then I will underline the object: $\underline{v}$.
Many texts use the notation: $\bar{v}$ or $\vec{v}$. I will avoid these notations due to the similarity, and possible confusion with, complex conjugation.

Representation I (Geometrically)

We will often visualize vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, and even in $\mathbb{R}^n$ as directed line segments. We assume that we can move the vectors around, as long as we do not change their magnitude and direction: that is, the vectors are not localized. Thus we can always move any vector so that its initial (or starting) point is any point in $\mathbb{R}^n$, and in particular, we may choose the initial (or starting point) to be the origin, O.

Note that in some physical problems, e.g. force, this may not be appropriate, as the point of application of the force is often important.



Representation II (Algebraically)

We write down our vectors as columns of numbers, these columns are often called n-**tuples** (especially when $n \geq 4$).

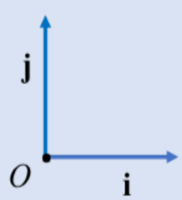
Example 2

$$\mathbf{w} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \in \mathbb{R}^2, \quad \mathbf{v} = \begin{pmatrix} -3 \\ 1 \\ -5 \end{pmatrix} \in \mathbb{R}^3, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n.$$

I am sure that most of you know exactly what is meant by these expressions. However there is actually more going on here than is apparent at first sight. A complete explanation will be given later on in the course.

At the present time **we should agree on the following: there are two special vectors in** $\mathbb{R}^2$: **i and j**, which we think of graphically as:

Diagram 2



When we write, for example, $\mathbf{w} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, what we mean is that $\mathbf{w} = 2\,\mathbf{i} + 3\,\mathbf{j}$.
The first version is a *shortand* for the second version.
For the first part of the course, we will use the basic vectors i and j, in $\mathbb{R}^2$, and **we say that we are using the *standard basis* in $\mathbb{R}^2$.**

Notation 2
We refer to the real numbers "2" and "3" as the **components** of the vector $\mathbf{w}$ in the standard basis in $\mathbb{R}^2$.

This interpretation is easily extended into higher dimensions. In $\mathbb{R}^3$ we have the three basic vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ and, in $\mathbb{R}^n$, we will indicate the standard basis by $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. So that

$$\mathbf{v} = \begin{pmatrix} 1 \\ -2 \\ 3 \\ -4 \\ 5 \end{pmatrix}$$

means

$$\mathbf{v} = 1\mathbf{e}_1 - 2\mathbf{e}_2 + 3\mathbf{e}_3 - 4\mathbf{e}_4 + 5\mathbf{e}_5.$$

Relation between the representations I and II

Let $\mathbf{p}$ be a vector in $\mathbb{R}^n$ which we may be thinking of as a directed line segment (type I representation). If we move that directed line segment so that its initial point is the origin, O, then we will label the terminal point with P (i.e., we use the same letter as the vector). If the point P has co-ordinates $(a_1, a_2, \ldots, a_n)$, then we will write the vector $\mathbf{p}$

as, $\mathbf{p} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$, that is, we equate this n-tuple with $\mathbf{p}$ (type II representation).

On the other hand, if $\mathbf{p}$ is a vector in $\mathbb{R}^n$ and we are thinking of $\mathbf{p}$ as an n-tuple,

$\mathbf{p} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ (type II representation), then we can draw a directed line segment from

(initial point) the origin, O, to (terminal point) the point P, with co-ordinates $(a_1, a_2, \ldots, a_n)$. We can now think of the vector $\mathbf{p}$ as this directed line segment (type I representation).

You are now free to move the vector $\mathbf{p}$ around, as long as you maintain its magnitude and direction, however, moving it will change both the initial and terminal points. ***We will say that the point P is the terminal point associated with the vector $\mathbf{p}$ (assuming that the initial point is O), or more simply, we say that P is the point associated with the vector $\mathbf{p}$. We will also say that the vector $\mathbf{p}$ is the vector associated with the terminal point P (assuming that the initial point is the origin O).***

Example 3 (continuation of Example 2)

The point W, with co-ordinates $(2, 3)$, is ***the (terminal) point associated with the vector $\mathbf{w}$, (assuming that the initial point is O)***. We can say more simply that W ***is the point in $\mathbb{R}^2$ associated with the vector $\mathbf{w}$***. We can also say that ***the vector $\mathbf{w}$ is the vector associated with the point*** W, with coordinates $(2, 3)$.

Similarly, the point V with coordinates $(-3, 1, -5)$, is the ***(terminal) point associated with the vector $\mathbf{v}$***, and the point X, with co-ordinates $(x_1, x_2, \ldots, x_n)$, is the **point in** $\mathbb{R}^n$ ***associated with the vector $\mathbf{x}$***.

Notation 3
When writing vectors out in component form we will always write them out as columns. This can be a nuisance at times as it takes up a lot of space, e.g.

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}.$$

You may find it more convenient and more compact to write vectors in a row. You can do so **as long as you also include the superscript** "T" (short for transpose) and you separate the components of the vectors with **commas**, as follows:

$$\mathbf{v} = (1, 2, 3, 4, 5)^T.$$

This means exactly the same as

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}.$$

Many texts do not make a distinction between row and column vectors, and they interchange between the two forms at random. This is careless and incorrect, we will see why later on. In this course, the notation presented in the lecture notes will have to be adhered to.

Remark : Equality of vectors in $\mathbb{R}^n$.

Two vectors $\mathbf{v} = (v_1, \ldots, v_n)^T$ and $\mathbf{w} = (w_1, \ldots, w_n)^T$ are equal, means:

(I) Geometrically: $\mathbf{v}$ and $\mathbf{w}$ have the same magnitude and same direction.

(II) Algebraically: their components are equal, that is, $v_k = w_k$, for all $k = 1, \ldots, n$.

Thus, the one vector equation $\mathbf{v} = \mathbf{w}$ is equivalent to n scalar equations, one for each component.

Operations on Vectors

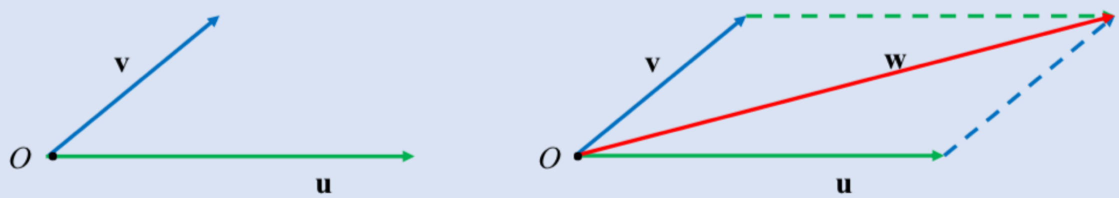
Definition 2: Addition

If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^n$, then we define their sum, $\mathbf{w}$, by $\mathbf{w} = \mathbf{u} + \mathbf{v}$.

(I) Geometrically

We can obtain **w** using the so-called parallelogram law, which we illustrate in $\mathbb{R}^3$:

Diagram 3



Move **v** without extending it or changing its direction, so that its starting point touches the terminal point of **u**. The new vector, **w**, is obtained by moving from the starting point of **u** to the terminal point of **v**. The vector **w** can also be obtained by moving **u** to the terminal point of **v**, creating thus a parallelogram.

Interpretation: suppose a boat is moving with a velocity of **u** and a person on the boat walks with velocity **v** on (relative to) the boat. If I am on the shore, I see the person move with velocity $\mathbf{w} = \mathbf{u} + \mathbf{v}$, and thus I see the combined velocity of the person and the boat.

(II) Algebraically

This is straightforwards, we obtain the components of the new vector $\mathbf{w} = \mathbf{u} + \mathbf{v}$ by adding the corresponding components of **u** and **v**.

Let $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix}$.

$$\text{Then,} \quad \mathbf{u} + \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_{n-1} + v_{n-1} \\ u_n + v_n \end{pmatrix}$$

$$\text{and} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n-1} \\ w_n \end{pmatrix} = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_{n-1} + v_{n-1} \\ u_n + v_n \end{pmatrix}, \text{ i.e., } w_i = u_i + v_i, \text{ for all } i = 1, 2, \ldots, n.$$

Example 4

$$\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} + \begin{pmatrix} 3 \\ -5 \\ 7 \end{pmatrix} = \begin{pmatrix} 2+3 \\ 4-5 \\ 6+7 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \\ 13 \end{pmatrix}.$$

The following properties of addition hold, and are stated for completion. They follow from the analogous statements about addition of real numbers. If I had not stated them then I suspect that you would have assumed that they were true anyway.

Lemma 1: Addition Rules

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$.

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

(ii) $\mathbf{u} + \mathbf{v} + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.

(iii) There is a special vector, $\mathbf{0}, \mathbf{0} = (0, 0, \ldots, 0, 0)^T$, in $\mathbb{R}^n$,
with the property that $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$. We call the vector $\mathbf{0}$, the **zero vector**.

Addition of m vectors is defined inductively, and via the Lemma 1. For example, to evaluate $\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4$ you can evaluate $\mathbf{v}_1 + \mathbf{v}_2$, and then add $\mathbf{v}_3$, and then add $\mathbf{v}_4$.

$$\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 = (\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3) + \mathbf{v}_4 = ((\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3) + \mathbf{v}_4.$$

Let $\mathbf{u} = (u_1, u_2, \ldots, u_{n-1}, u_n)^T \in \mathbb{R}^n$. Let us consider a vector having the same magnitude as $\mathbf{u}$, but having opposite direction. We denote this vector by $-\mathbf{u}$, and observe then that

$$-\mathbf{u} = (-u_1, -u_2, \ldots, -u_{n-1}, -u_n)^T.$$

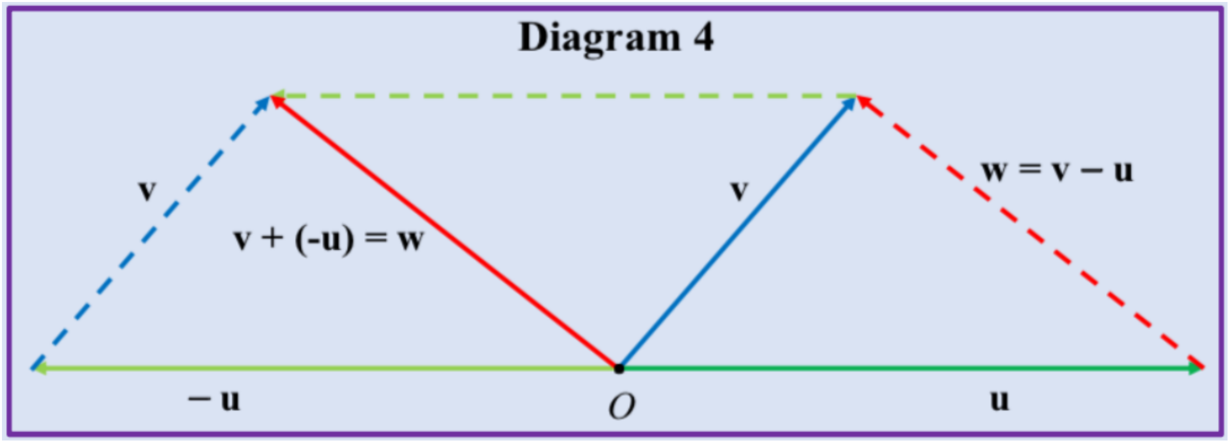
Definition 3: Subtraction

Let $\mathbf{v} = (v_1, v_2, \ldots, v_{n-1}, v_n)^T$, $\mathbf{u} = (u_1, u_2, \ldots, u_{n-1}, u_n)^T$ and $-\mathbf{u} = (-u_1, -u_2, \ldots, -u_{n-1}, -u_n)^T$ be vectors in $\mathbb{R}^n$. We then to define subtraction as follows:

$$\mathbf{v} - \mathbf{u} \ \text{ means } \mathbf{v} + (-\mathbf{u}).$$

So that

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix} - \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix} + \begin{pmatrix} -u_1 \\ -u_2 \\ \vdots \\ -u_{n-1} \\ -u_n \end{pmatrix} = \begin{pmatrix} v_1 - u_1 \\ v_2 - u_2 \\ \vdots \\ v_{n-1} - u_{n-1} \\ v_n - u_n \end{pmatrix}.$$



Example 5

$$\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} - \begin{pmatrix} 3 \\ -5 \\ 7 \end{pmatrix} = \begin{pmatrix} 2-3 \\ 4+5 \\ 6-7 \end{pmatrix} = \begin{pmatrix} -1 \\ 9 \\ -1 \end{pmatrix}.$$

Lemma 2: Cancellation Identity

Let $\mathbf{v} \in \mathbb{R}^n$. Then

$$\mathbf{v} - \mathbf{v} = \mathbf{v} + (-\mathbf{v}) = (-\mathbf{v}) + \mathbf{v} = \mathbf{0}.$$

The vector $-\mathbf{v}$ has the effect of cancelling the vector $\mathbf{v}$ when addition is performed, and so $-\mathbf{v}$ is called the **additive inverse** of $\mathbf{v}$. Adding one and then the other is the same as adding zero, and so has no effect!

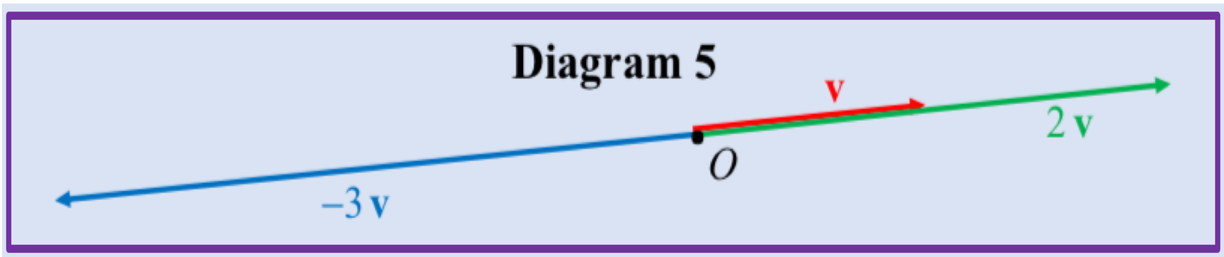
Note that **we do not have an operation to multiply vectors together** (in $\mathbb{R}^n$, with $n > 1$). However we can multiply a vector by a real number. This process is often called scaling the vector, and the real numbers can be referred to as **scalars**.

Definition 4: Scalar Multiplication

Let $p \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$ with $\mathbf{v} = (v_1, v_2, \ldots, v_{n-1}, v_n)^T$.

We define $p\mathbf{v}$ by:
$$p\mathbf{v} = p \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix} = \begin{pmatrix} pv_1 \\ pv_2 \\ \vdots \\ pv_{n-1} \\ pv_n \end{pmatrix}.$$

Each of the components of $\mathbf{v}$ are multiplied by the scalar p.
The vector $\mathbf{v}$ has been scaled by the real number (scalar) p.



The vector $p\mathbf{v}$ points in the direction of $\mathbf{v}$ (or $-\mathbf{v}$) and is $|p|$ times as long. Notice when $p < 0$, the new vector, $p\mathbf{v}$, points in the opposite direction to $\mathbf{v}$. The vector $(-1)\mathbf{v} = -\mathbf{v}$, has thus the same length as $\mathbf{v}$, but points in the opposite direction to $\mathbf{v}$.

Here is another result that I am stating for completeness. Were it not stated, I am sure that you all would have used these results anyway, since they are very natural.

Lemma 3 : Some properties of scalar multiplication

Let $\mathbf{u}$ and $\mathbf{v}$ be in $\mathbb{R}^n$, and p and q be real numbers.

(ii) $(p+q)\,\mathbf{v} = p\,\mathbf{v} + q\,\mathbf{v}$.

(ii) $(p\,q)\mathbf{v} = p\,(q\mathbf{v})$.

(iii) $p\,(\mathbf{u}+\mathbf{v}) = p\,\mathbf{u} + p\,\mathbf{v}$.

(iv) $0\,\mathbf{v} = \mathbf{0}$.

The last result that we state in Topic 1 is called **properties of zero**. You have used the following result often in mathematics when you have been dealing with real (or complex) equations:

Let $a, b \in \mathbb{R}$, if $ab = 0$, then either $a = 0$ or $b = 0$.

There is an analogous result for vectors, and we will use it often when we are solving vector equations.

Lemma 4: Properties of zero (Cancellation Law).

Let $a \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$.

If $a\mathbf{v} = \mathbf{0}$, then either $a = 0$ or $\mathbf{v} = \mathbf{0}$.

Vectors in $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^n$ are used widely in mechanics for displacement, velocity, acceleration, force, and many other important physical concepts. We will not be dwelling too much on these applications in this course.

$\mathbb{C}^n$

Although we do not have the same geometric interpretation, we introduce $\mathbb{C}^n$ via columns of n complex numbers.

Definition 5: $\mathbb{C}^n$ and complex n-vectors

We define $\mathbb{C}^n = \left\{ \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} : z_i \in \mathbb{C}, i = 1, 2, \cdots, n \right\}$ and refer to $\mathbf{z} \in \mathbb{C}^n$ by a **complex n-vector**, or just a complex vector when n is clear.

Example 6

$$\mathbf{z} = \begin{pmatrix} 1+i \\ 2-3i \\ -4+5i \end{pmatrix} \in \mathbb{C}^3$$

$$\mathbf{w} = \begin{pmatrix} 1-2i \\ 7 \end{pmatrix} \in \mathbb{C}^2.$$

Algebraic equality, addition and subtraction of elements of $\mathbb{C}^n$ are defined in the obvious way and satisfy Lemma 1 and Lemma 2, with $\mathbb{C}$ replacing $\mathbb{R}$. Scalar multiplication of elements of $\mathbb{C}^n$ by elements in $\mathbb{C}$ is defined in the obvious way, i.e. component-wise and satisfies Lemma 3. Finally, Lemma 4 holds in $\mathbb{C}^n$ too.

Example 7

$$3\begin{pmatrix} 1+i \\ 2+i \end{pmatrix} - 2\begin{pmatrix} 2-i \\ 4+i \end{pmatrix} = \begin{pmatrix} 3+3i \\ 6+3i \end{pmatrix} - \begin{pmatrix} 4-2i \\ 8+2i \end{pmatrix} = \begin{pmatrix} -1+5i \\ -2+i \end{pmatrix}.$$

Topic 2

The Dot Product in $\mathbb{R}^n$.

The dot product is an operation which converts two vectors in $\mathbb{R}^n$ into a real number.

Definition 1: Dot product

Let $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n-1} \\ w_n \end{pmatrix}$ be two vectors in $\mathbb{R}^n$, then we define their

dot product, $\mathbf{v} \bullet \mathbf{w}$, to mean the real number obtained by multiplying the corresponding components together and then adding the resulting terms up:

$$\mathbf{v} \bullet \mathbf{w} = v_1 w_1 + v_2 w_2 + \cdots + v_{n-1} w_{n-1} + v_n w_n.$$

Examples 1

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} \bullet \begin{pmatrix} 5 \\ 7 \end{pmatrix} = 2(5) + 3(7) = 31$$

$$\begin{pmatrix} 3 \\ -5 \\ 2 \end{pmatrix} \bullet \begin{pmatrix} -4 \\ 4 \\ -2 \end{pmatrix} = (3(-4)) + ((-5)4) + (2(-2)) = -36.$$

There are some very important properties which the dot product naturally satisfies.

Lemma 1: Properties of the dot product

If $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{z}$ are vectors in $\mathbb{R}^n$ and $a \in \mathbb{R}$, then

(i) $\mathbf{v} \bullet \mathbf{w} = \mathbf{w} \bullet \mathbf{v}$ (called symmetry).

$$\left.\begin{array}{l} \text{(ii) } (\mathbf{v} + \mathbf{w}) \bullet \mathbf{z} = \mathbf{v} \bullet \mathbf{z} + \mathbf{w} \bullet \mathbf{z} \\ \\ \text{(iii) } (a\mathbf{w}) \bullet \mathbf{v} = a(\mathbf{w} \bullet \mathbf{v}) \end{array}\right\} \text{(collectively known as linearity).}$$

(iv) $\mathbf{v} \bullet \mathbf{v} \geq 0$, with equality iff $\mathbf{v} = \mathbf{0}$ (non-negativity).

Definition 2: The Argument of a function

The argument of a function is another name for the independent variable. In real single variable Calculus, you might have $y = f(x)$ and so we could refer to the real independent variable "x" as the argument of the function f.

The dot product can be thought of as a function with two arguments which are both vectors in $\mathbb{R}^n$. We would write this as:

$$< \ , \ >: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad \text{defined by} \quad < \mathbf{x}\, , \mathbf{y} >= \mathbf{x} \bullet \mathbf{y}.$$

We can refer to "$\mathbf{x}$" as the first argument and we can refer to "$\mathbf{y}$" as the second argument.

In Lemma 1, properties (ii) and (iii) are often stated in words as, "the dot product is linear in its first argument." It may also be shown that the dot product is linear in its second argument too.

One of the major uses of the dot products is in defining lengths of vectors.

Definition 3: Length

The **length** (or **norm**) of the vector $\mathbf{v} \in \mathbb{R}^n$ is denoted by $\|\mathbf{v}\|$ and is obtained by perfoming the calculation:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \bullet \mathbf{v}}.$$

Example 2

$$\|\begin{pmatrix} 2 \\ 4 \end{pmatrix}\| = \sqrt{\begin{pmatrix} 2 \\ 4 \end{pmatrix} \bullet \begin{pmatrix} 2 \\ 4 \end{pmatrix}} = \sqrt{20}$$

$$\|\begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}\| = \sqrt{\begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} \bullet \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}} = \sqrt{29}.$$

The notation "$\|$" is used so that you do not get confused with "$|$" used for the absolute value of a real number.

Recall that when we take the square root in mathematics, we do mean the positive square root, so that $\sqrt{4} = 2$.

Notice that it follows from Lemma 1 (iv) that every vector has a length greater than zero, except the zero vector itself which has a length of zero.

Lemma 2

If $\mathbf{v} \in \mathbb{R}^n$ and if $a \in \mathbb{R}$, then
$$\|a\mathbf{v}\| = |a|\ \|\mathbf{v}\|.$$

Examples 3

$$\|\begin{pmatrix} 3 \\ 9 \end{pmatrix}\| = \|3\begin{pmatrix} 1 \\ 3 \end{pmatrix}\| = |3|\ \|\begin{pmatrix} 1 \\ 3 \end{pmatrix}\| = 3\sqrt{10}$$

$$\|\begin{pmatrix} -2 \\ -6 \\ 4 \end{pmatrix}\| = \|(-2)\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}\| = |-2|\ \|\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}\| = 2\sqrt{14}.$$

Definition 4 : Unit vector

We say that $\mathbf{w} \in \mathbb{R}^n$ is a **unit vector** to mean that $\|\mathbf{w}\| = 1$.

Examples 4

$$\text{If } \mathbf{u} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \text{ then } \|\mathbf{u}\| = \sqrt{13} \text{ and so } \mathbf{u} \text{ is not a unit vector.}$$

$$\text{If } \mathbf{v} = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)^T, \text{ then } \|\mathbf{v}\| = 1 \text{ and so } \mathbf{v} \text{ is a unit vector.}$$

When $\mathbf{z} \in \mathbb{R}^n$ is a non-zero vector, we can produce a unit vector in the direction of $\mathbf{z}$, which we denote by $\hat{\mathbf{z}}$, by scaling it. This process is called **normalization**.
$$\hat{\mathbf{z}} = \frac{\mathbf{z}}{\|\mathbf{z}\|}.$$
The vector $\hat{\mathbf{z}}$ is a unit vector since
$$\|\hat{\mathbf{z}}\| = \|\frac{\mathbf{z}}{\|\mathbf{z}\|}\| = \frac{\|\mathbf{z}\|}{\|\mathbf{z}\|} = 1.$$

The vector $\hat{\mathbf{z}}$ has the same direction of the vector $\mathbf{z}$ since
$$\hat{\mathbf{z}} = \frac{\mathbf{z}}{\|\mathbf{z}\|} = \frac{1}{\|\mathbf{z}\|}\,\mathbf{z}.$$

Example 5

We produce a unit vector in the direction of $\mathbf{w} = \begin{pmatrix} 2 \\ -4 \\ 6 \end{pmatrix}$, that is the vector

$$\hat{\mathbf{w}} = \frac{\mathbf{w}}{\|\mathbf{w}\|} = \frac{1}{\sqrt{56}} \begin{pmatrix} 2 \\ -4 \\ 6 \end{pmatrix}.$$

Definition 5: Orthogonal

We say that the two vectors $\mathbf{v}$ and $\mathbf{w}$ (in $\mathbb{R}^n$) are **orthogonal** to mean that $\mathbf{v} \bullet \mathbf{w} = 0$.

Examples 6

$$\mathbf{v} = \begin{pmatrix} 1 \\ -3 \\ 5 \end{pmatrix} \text{ and } \mathbf{w} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \text{ are orthogonal as } \begin{pmatrix} 1 \\ -3 \\ 5 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = 0.$$

$$\mathbf{v} = \begin{pmatrix} 1 \\ -3 \\ 5 \end{pmatrix} \text{ and } \mathbf{u} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \text{ are not orthogonal as } \begin{pmatrix} 1 \\ -3 \\ 5 \end{pmatrix} \bullet \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} = 1 \neq 0.$$

Note that every vector $\mathbf{v} \in \mathbb{R}^n$ is orthogonal to the zero vector, that is:

$$\mathbf{v} \bullet \mathbf{0} = v_1 (0) + \cdots + v_n (0) = 0.$$

If you have a single vector, then it is easy to produce one that is orthogonal to it:

For instance:

$$\begin{pmatrix} -b \\ a \end{pmatrix} \text{ is orthogonal to } \begin{pmatrix} a \\ b \end{pmatrix} \text{ in } \mathbb{R}^2, \text{ since } \begin{pmatrix} -b \\ a \end{pmatrix} \bullet \begin{pmatrix} a \\ b \end{pmatrix} = 0$$

and

$$\begin{pmatrix} -b \\ a \\ 0 \end{pmatrix} \text{ is orthogonal to } \begin{pmatrix} a \\ b \\ c \end{pmatrix} \text{ in } \mathbb{R}^3, \text{ since } \begin{pmatrix} -b \\ a \\ 0 \end{pmatrix} \bullet \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0.$$

Can you produce other vectors which are orthogonal to $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$?

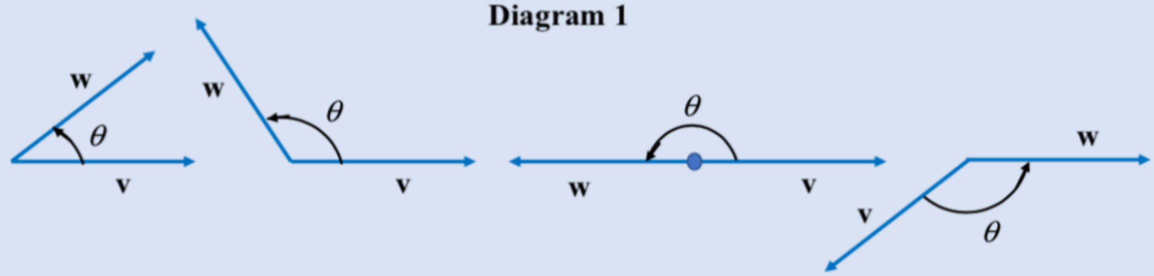
The Dot Product And Geometry

Definition 6: Angle

Let $\mathbf{v}$ and $\mathbf{w}$ be two non-zero vectors in $\mathbb{R}^n$ then we can define the **angle,** in radians, θ $(0 \leq \theta \leq \pi)$, between these two vectors as follows:

$$\mathbf{v} \bullet \mathbf{w} = \|\mathbf{v}\| \, \|\mathbf{w}\| \cos\theta, \quad \text{i.e.} \quad \theta = \arccos\left(\frac{\mathbf{v} \bullet \mathbf{w}}{\|\mathbf{v}\| \, \|\mathbf{w}\|}\right).$$

Diagram 1



The angle θ between $\mathbf{v}$ and $\mathbf{w}$ is between 0 and π

Examples 7

Find the angle θ between $\begin{pmatrix} 1 \\ 4 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 3 \end{pmatrix}$ in $\mathbb{R}^2$.

Solution - Note that we will work to three decimal places throughout.

$$\theta = \arccos\left(\frac{\begin{pmatrix} 1 \\ 4 \end{pmatrix} \bullet \begin{pmatrix} -2 \\ 3 \end{pmatrix}}{\| \begin{pmatrix} 1 \\ 4 \end{pmatrix} \| \, \| \begin{pmatrix} -2 \\ 3 \end{pmatrix} \|}\right) = \arccos\left(\frac{10}{\sqrt{(17)(13)}}\right) = \arccos(0.672) = 0.833.$$

Find the angle, φ, between $\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ -4 \\ 3 \end{pmatrix}$ in $\mathbb{R}^3$.

Solution - Note that we will work to three decimal places throughout.

$$\varphi = \arccos\left(\frac{\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} \bullet \begin{pmatrix} 2 \\ -4 \\ 3 \end{pmatrix}}{\| \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} \| \, \| \begin{pmatrix} 2 \\ -4 \\ 3 \end{pmatrix} \|}\right) = \arccos\left(\frac{5}{\sqrt{(35)(29)}}\right) = \arccos(0.157) = 1.413.$$

Notice that this definition of angle between two vectors is consistent with our previous knowledge of orthogonality in that, if the two vectors are orthogonal, then their dot product is zero, and we deduce that the angle θ between them is $\frac{\pi}{2}$, since $\cos\frac{\pi}{2} = 0$.

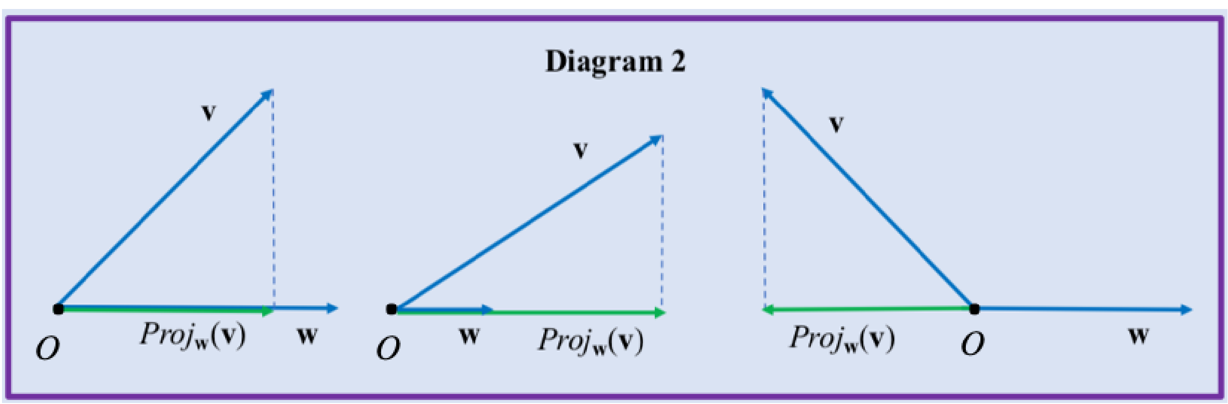
The Dot Product, Projection and Remainder.

Suppose that $\mathbf{w}$ is a non-zero vector. Then we can think of projecting any vector $\mathbf{v}$ (even $\mathbf{w}$ itself) along $\mathbf{w}$, or equivalently, project $\mathbf{v}$ onto $\mathbf{w}$.

Definition 7: The projection of $\mathbf{v}$ along $\mathbf{w}$.

Let $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ with $\mathbf{w} \neq \mathbf{0}$.
The **projection of v along w**, or the **projection of v in the w direction,** $Proj_{\mathbf{w}}(\mathbf{v})$, is defined by:

$$Proj_{\mathbf{w}}(\mathbf{v}) = \frac{(\mathbf{v} \bullet \mathbf{w})}{\|\mathbf{w}\|^2}\mathbf{w}.$$



Example 8

Find the projection of $\mathbf{u} = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}$ in the direction of $\mathbf{w} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Solution

$$Proj_{\mathbf{w}}(\mathbf{u}) = \frac{(\mathbf{u} \bullet \mathbf{w})}{\|\mathbf{w}\|^2}\mathbf{w} = \frac{\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}}{\|\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}\|^2} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{10}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{5}{7}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Notice that we can write the projection as follows:

$$Proj_{\mathbf{w}}(\mathbf{v}) = \frac{(\mathbf{v} \bullet \mathbf{w})}{\|\mathbf{w}\|^2}\,\mathbf{w} = \frac{(\mathbf{v} \bullet \mathbf{w})}{\|\mathbf{w}\|}\,\frac{\mathbf{w}}{\|\mathbf{w}\|} = (\mathbf{v} \bullet \hat{\mathbf{w}})\,\hat{\mathbf{w}}.$$

This last expression clearly shows that it is the **direction** of the vector $\mathbf{w}$, indicated by the unit vector $\hat{\mathbf{w}}$, that is important, rather than the actual vector $\mathbf{w}$ itself.
The **magnitude of $\mathbf{w}$ is not relevant** to the projection, but the direction of $\mathbf{w}$ is.

If the angle between $\mathbf{v}$ and $\mathbf{w}$ is θ, then we can also write:

$$Proj_{\mathbf{w}}(\mathbf{v}) = (\|\mathbf{v}\|\cos(\theta))\hat{\mathbf{w}}.$$

We think of the projection of $\mathbf{v}$ along $\mathbf{w}$ as giving that part of the vector $\mathbf{v}$ that lies in the $\mathbf{w}$ direction. And, in fact, we call the scalar factor of this unit vector, the component.

Definition 8: Component

Let $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ with $\mathbf{w} \neq \mathbf{0}$.

We refer to the quantity

$$\|\mathbf{v}\|\cos(\theta),$$

as the **component** of $\mathbf{v}$ along $\mathbf{w}$. Since this quantity is a scalar, it is also called the **scalar component** of $\mathbf{v}$ in the $\mathbf{w}$ direction.

If the scalar component is negative, then the projection of $\mathbf{v}$ along $\mathbf{w}$ will be in the direction of the vector $-\mathbf{w}$. It is traditional to still think of this projection vector as being in the $\mathbf{w}$ direction.

Example 9

Find the component of $\mathbf{u} = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}$ in the $\mathbf{w}$ direction, where $\mathbf{w} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Solution

We have done the calculation in Example 8:

$$Proj_{\mathbf{w}}(\mathbf{u}) = \frac{5}{7}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{5}{7}\sqrt{14}\left[\frac{1}{\sqrt{14}}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}\right].$$

Thus the component of $\mathbf{u}$ in the $\mathbf{w}$ direction is $\frac{5}{7}\sqrt{14}$.

Definition 9: Remainder

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ with $\mathbf{w} \neq \mathbf{0}$.

We refer to the vector $\mathbf{v} - Proj_{\mathbf{w}}(\mathbf{v})$ as the **remainder**, and denote it by $\mathbf{r}$,

$$\mathbf{r} = \mathbf{v} - Proj_{\mathbf{w}}(\mathbf{v}).$$

Example 10

What is the remainder when $\mathbf{u} = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}$ is projected onto $\mathbf{w} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$?

Solution

$$\mathbf{r} = \mathbf{u} - Proj_{\mathbf{w}}(\mathbf{u}) = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} - \frac{5}{7}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{7}\begin{pmatrix} 16 \\ -38 \\ 20 \end{pmatrix}.$$

Lemma 3

The projection of a vector $\mathbf{v}$ along $\mathbf{w}$ and the remainder are orthogonal to each other.

Proof

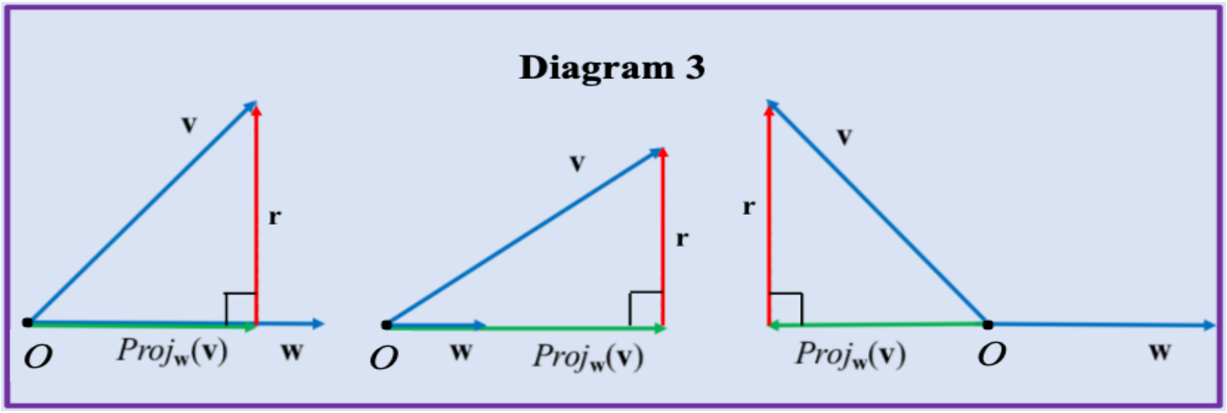
This remainder is orthogonal to $\mathbf{w}$, since

$$\mathbf{r} \bullet \mathbf{w} = (\mathbf{v} - Proj_{\mathbf{w}}(\mathbf{v})) \bullet \mathbf{w} = \mathbf{v} \bullet \mathbf{w} - (Proj_{\mathbf{w}}(\mathbf{v})) \bullet \mathbf{w}$$

$$= \mathbf{v} \bullet \mathbf{w} - \left(\frac{(\mathbf{v} \bullet \mathbf{w})}{||\mathbf{w}||}\frac{\mathbf{w}}{||\mathbf{w}||}\right) \bullet \mathbf{w} = \mathbf{v} \bullet \mathbf{w} - \left(\frac{(\mathbf{v} \bullet \mathbf{w})}{||\mathbf{w}||}\right)\frac{\mathbf{w} \bullet \mathbf{w}}{||\mathbf{w}||}$$

$$= \mathbf{v} \bullet \mathbf{w} - \left(\frac{(\mathbf{v} \bullet \mathbf{w})}{||\mathbf{w}||}\right)\frac{||\mathbf{w}||^2}{||\mathbf{w}||}) = \mathbf{v} \bullet \mathbf{w} - \mathbf{v} \bullet \mathbf{w} = 0. \qquad \blacksquare$$

Since the remainder is orthogonal to $\mathbf{w}$ it is also orthogonal to the projection of $\mathbf{v}$ onto $\mathbf{w}$. Thus $\mathbf{r}$ is a vector which lies in the space that is orthogonal to $\mathbf{w}$. We think of $\mathbf{r}$ as the projection of $\mathbf{v}$ onto the space (hyper-plane) orthogonal to $\mathbf{w}$. Some texts write this as $Perp_{\mathbf{w}}(\mathbf{v})$, read as "perp of $\mathbf{v}$ onto $\mathbf{w}$."



$$\mathbf{r} = \mathbf{v} - Proj_{\mathbf{w}}(\mathbf{v}) \quad \text{or equivalently} \quad \mathbf{v} = \mathbf{r} + Proj_{\mathbf{w}}(\mathbf{v})$$

Topic 3

The Standard Inner Product on $\mathbb{C}^n$.

We are already aware that there is a difference in finding the magnitudes of real numbers and in finding the magnitudes of complex numbers:

$$\text{if } x \in \mathbb{R}, \text{ then the length of } x \text{ is } |x| = \sqrt{x(x)},$$

$$\text{if } z \in \mathbb{C}, \text{ then the length (modulus) of } z \text{ is } |z| = \sqrt{z(\bar{z})}.$$

Note that we have to introduce the complex conjugate in this second expression. If we do not, and just have $z(z)$ instead, then we have a complex number. The square root of that complex number, which is two complex numbers, would not tell us about the length of the original complex number z.

Example 1

What is the length of $w = 2 - 3i$?

Solution

$$|w| = |2 - 3i| = \sqrt{(2 - 3i)\,(\,\overline{2 - 3i}\,)} = \sqrt{2^2 + (-3)^2} = \sqrt{13}.$$

One of the most important uses of the dot product on $\mathbb{R}^n$ is in finding the lengths of vectors. When we try to extend this concept to $\mathbb{C}^n$, we will have the same issue as above, that is, we will need to introduce complex conjugation at one point.

Definition 1: The Standard Inner Product on $\mathbb{C}^n$.

Let $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{n-1} \\ w_n \end{pmatrix}$ and $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ z_n \end{pmatrix}$ be two vectors in $\mathbb{C}^n$. We then define their standard inner product $\langle \mathbf{w}, \mathbf{z} \rangle$, to mean the following quantity:

$$\langle \mathbf{w}, \mathbf{z} \rangle = w_1\overline{z}_1 + w_2\overline{z}_2 + \cdots + w_{n-1}\overline{z}_{n-1} + w_n\overline{z}_n.$$

Notice that this is similar to the dot product, in that we multiply terms together and add them up. However there is a major important difference, and that is the fact that we take the complex conjugate of all of the components that come from the second argument, in this case $\mathbf{z}$.

Example 2

Evaluate $\left\langle \begin{pmatrix} 2-3i \\ 1-2i \end{pmatrix}, \begin{pmatrix} -3+4i \\ 3+5i \end{pmatrix} \right\rangle$.

Solution

$$\begin{aligned} \left\langle \begin{pmatrix} 2-3i \\ 1-2i \end{pmatrix}, \begin{pmatrix} -3+4i \\ 3+5i \end{pmatrix} \right\rangle &= (2-3i)\left(\overline{-3+4i}\right) + (1-2i)\left(\overline{3+5i}\right) \\ &= (2-3i)(-3-4i) + (1-2i)(3-5i) \\ &= -6-8i+9i-12+3-5i-6i-10 \\ &= -25-10i \end{aligned}$$

Lemma 1: Properties of the standard inner product on $\mathbb{C}^n$

If $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{z}$ are vectors in $\mathbb{C}^n$ and $a \in \mathbb{C}$, then

(i) $\langle \mathbf{v}, \mathbf{z} \rangle = \overline{\langle \mathbf{z}, \mathbf{v} \rangle}$ (CONJUGATE symmetry)

$$\left.\begin{array}{l} \text{(ii)} \ \langle (\mathbf{v}+\mathbf{w}), \mathbf{z} \rangle = \langle \mathbf{v}, \mathbf{z} \rangle + \langle \mathbf{w}, \mathbf{z} \rangle \\ \text{(iii)} \ \langle a\mathbf{v}, \mathbf{w} \rangle = a \langle \mathbf{v}, \mathbf{w} \rangle \end{array}\right\} \text{(linearity in the first argument).}$$

(iv) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, with equality **iff** $\mathbf{v} = \mathbf{0}$ (non-negativity).

As in the real case, we can make use of the non-negativity property of Lemma 1, i.e.

(iv) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, with equality iff $\mathbf{v} = \mathbf{0}$,

to define lengths of vectors, for vectors now in $\mathbb{C}^n$.

Definition 2: Lengths of vectors in $\mathbb{C}^n$.

Let $\mathbf{v} \in \mathbb{C}^n$, we define the **length** of $\mathbf{v}$, denoted by $\|\mathbf{v}\|$, as $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

Example 3

What is the length of the vector $\mathbf{v} = \begin{pmatrix} 2-i \\ -3+2i \\ -4-5i \end{pmatrix}$?

Solution

$$\begin{aligned}\langle \mathbf{v}, \mathbf{v} \rangle &= \left\langle \begin{array}{c} 2-i \\ -3+2i \\ -4-5i \end{array}, \begin{array}{c} 2-i \\ -3+2i \\ -4-5i \end{array} \right\rangle \\ &= (2-i)(\overline{2-i}) + (-3+2i)(\overline{-3+2i}) + (-4-5i)(\overline{-4-5i}) \\ &= 2^2+1^2+3^2+2^2+4^2+5^2 = 59.\end{aligned}$$

We conclude that $\mathbf{v}$ has length of $\sqrt{59}$.

Note that

$$\langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2 = |2-i|^2 + |-3+2i|^2 + |-4-5i|^2.$$

Lemma 2: Properties of the length.

Let $\mathbf{v} \in \mathbb{C}^n$ and $c \in \mathbb{C}$, then:

(i) $\|c\mathbf{v}\| = |c|\,\|\mathbf{v}\|$

(ii) $\|\mathbf{v}\| \geq 0$, with equality **iff** $\mathbf{v} = \mathbf{0}$.

Example 4

What is $\|(1+2i)\begin{pmatrix} 2-i \\ -3+2i \\ -4-5i \end{pmatrix}\|$?

Solution (using Example 3)

$$\|(1+2i)\begin{pmatrix} 2-i \\ -3+2i \\ -4-5i \end{pmatrix}\| = |1+2i|\,\|\begin{pmatrix} 2-i \\ -3+2i \\ -4-5i \end{pmatrix}\| = \sqrt{5}\,\sqrt{59} = \sqrt{295}.$$

Definition 3: Orthogonality in $\mathbb{C}^n$

Let $\mathbf{v}$ and $\mathbf{w}$ be two vectors in $\mathbb{C}^n$.

We say that $\mathbf{v}$ and $\mathbf{w}$ are orthogonal to mean that $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

Example 5

Let $\mathbf{v} = \begin{pmatrix} 1+i \\ 1+2i \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} 2+i \\ 1+i \end{pmatrix}$, and $\mathbf{z} = \begin{pmatrix} 2+i \\ -1-i \end{pmatrix}$.

Evaluate $\langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle \mathbf{v}, \mathbf{z} \rangle$.

Solution

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= (1+i)(\overline{2+i}) + (1+2i)(\overline{1+i}) \\ &= 2 - i + 2i + 1 + 1 - i + 2i + 2 \\ &= 6 + 2i. \end{aligned}$$

and

$$\begin{aligned} \langle \mathbf{v}, \mathbf{z} \rangle &= (1+i)(\overline{2+i}) + (1+2i)(\overline{-1-i}) \\ &= 2 - i + 2i + 1 - 1 - 2i + i - 2 \\ &= 0. \end{aligned}$$

Thus, $\mathbf{v}$ and $\mathbf{z}$ are orthogonal.

Example 6

Find a unit vector that is orthogonal to the vector $\mathbf{w} = \begin{pmatrix} 2+i \\ 1+i \end{pmatrix}$ in Example 5.

Solution

We are looking for a vector $\mathbf{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ satisfying 2 conditions:

$\langle \mathbf{x}, \mathbf{x} \rangle = 1$, so that

$$\alpha\bar{\alpha} + \beta\bar{\beta} = 1 \qquad (*)$$

and

$\langle \mathbf{x}, \mathbf{w} \rangle = 0$, so that

$$\alpha(2-i) + \beta(1-i) = 0. \qquad (**)$$

If we let $\alpha = a + ci$ and $\beta = b + di$, with a, b, c, d all in $\mathbb{R}$, then we have that:

$$(*) \Longrightarrow a^2 + c^2 + b^2 + d^2 = 1$$

and

$$(**) \Longrightarrow (a+ic)(2-i) + (b+id)(1-i) = 0.$$

This is a system of two equations for four unknowns, and thus there is lots of freedom.

I **choose** to find a solution by letting $c = 0$. We then have:

$$(*) \Longrightarrow a^2 + b^2 + d^2 = 1$$

and

$$(**) \Longrightarrow a(2-i) + b(1-i) + id(1-i) = 0.$$

Considering the real and imaginary parts of $(**)$, we get:

$$2a + b + d = 0 \ \text{ and } \ -a - b + d = 0.$$

Adding these last equations gives that

$$a + 2d = 0, \ \text{ so that } \ a = -2d, \ \text{ and then, } \ b = -(2a+d) = 3d.$$

Finally, $(*)$ is then

$$a^2 + b^2 + d^2 = 1 \Longrightarrow (4+9+1)d^2 = 1 \ \Longrightarrow d^2 = \frac{1}{14}.$$

A solution is $d = \dfrac{1}{\sqrt{14}}$, and then

$$\alpha = a = -\frac{2}{\sqrt{14}} \ \text{ and } \ \beta = b + id = \frac{3}{\sqrt{14}} + \frac{1}{\sqrt{14}} i,$$

and thus, a solution to the original problem is:

$$\mathbf{x} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{\sqrt{14}} \begin{pmatrix} -2 \\ 3+i \end{pmatrix}.$$

Definition 4: The Projection of $\mathbf{z}$ along $\mathbf{w}$

Let $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$, with $\mathbf{w} \neq \mathbf{0}$.

The **projection of z along w** , or the projection of $\mathbf{z}$ in the $\mathbf{w}$ direction, $Proj_{\mathbf{w}}(\mathbf{z})$, is defined by

$$Proj_{\mathbf{w}}(\mathbf{z}) = \frac{\langle \mathbf{z}, \mathbf{w} \rangle \, \mathbf{w}}{\|\mathbf{w}\|^2} = \langle \mathbf{z}, \hat{\mathbf{w}} \rangle \hat{\mathbf{w}}.$$

Example 7

Find the projection of $\mathbf{z} = \begin{pmatrix} 1 \\ i \\ 1+i \end{pmatrix}$ in the direction of $\mathbf{w} = \begin{pmatrix} 1-i \\ 2-i \\ 3+i \end{pmatrix}$.

Solution

$$
\begin{aligned}
\|\mathbf{w}\|^2 &= |\,1-i\,|^2 + |\,2-i\,|^2 + |\,3+i\,|^2 \\
&= 1+1+4+1+9+1 = 17
\end{aligned}
$$

$$
\begin{aligned}
\langle \mathbf{z},\, \mathbf{w} \rangle &= 1(1+i) + i(2+i) + (1+i)(3-i) \\
&= 1 + i + 2i - 1 + 3 - i + 3i + 1 \\
&= 4 + 5i.
\end{aligned}
$$

The solution is thus:

$$
Proj_{\mathbf{w}}(\mathbf{z}) = \frac{4+5i}{17} \begin{pmatrix} 1-i \\ 2-i \\ 3+i \end{pmatrix}.
$$

Field

During the course, we will be working both with the universal set of $\mathbb{R}$, and vectors in $\mathbb{R}^n$, and also with the universal set of $\mathbb{C}$, and vectors in $\mathbb{C}^n$.

In much of what we do, there is no need to make an explicit distinction, and we will refer to our universal set as the field $\mathbb{F}$. The field $\mathbb{F}$ will either be $\mathbb{R}$ or $\mathbb{C}$, and most of the time, we do not need to declare which one it is. There are a few occasions in MATH 136 where we need to be explicit about which field we are using and there will be many occasions in MATH 235 where we need to tell you what the field is.

For example, if $x \in \mathbb{F}$, then solve $2x = a$.

The solution is $x = \dfrac{a}{2}$ (the exact field was not needed).

But if $x \in \mathbb{F}$, then solve $x^2 = -1$, we need to know exactly what $\mathbb{F}$ is.

(a) If $\mathbb{F} = \mathbb{R}$, then this equation has no solutions.

(b) If $\mathbb{F} = \mathbb{C}$, then $x = \pm i$.

We will refer to, $x \in \mathbb{F}$, as being a **scalar**.

Definition 5: The Standard Inner Product on $\mathbb{F}^n$

Let $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, then we define the **standard inner product on** $\mathbb{F}^n$ by:

$$
\langle \mathbf{x}, \mathbf{y} \rangle = x_1\bar{y}_1 + x_2\bar{y}_2 + \cdots + x_n\bar{y}_n.
$$

Notice that:

If $\mathbb{F} = \mathbb{R}$, then this inner product is the dot product on $\mathbb{R}^n$.

If $\mathbb{F} = \mathbb{C}$, then this inner product is the standard inner product on $\mathbb{C}^n$.

Topic 4
The Cross Product

This is ONLY defined for $\mathbb{R}^3$

Definition 1: The cross product

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^3$, with, $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, then we define the **cross product** of $\mathbf{u}$ and $\mathbf{v}$, written as $\mathbf{u} \times \mathbf{v}$, to be the vector in $\mathbb{R}^3, \mathbf{z}$, given by

$$\mathbf{z} = \mathbf{u} \times \mathbf{v} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ -[u_1v_3 - u_3v_1] \\ u_1v_2 - u_2v_1 \end{pmatrix}.$$

Example 1

$$\begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} \times \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} (-3)(4) - (5)(1) \\ -((2)(4) - (5)(-2)) \\ (2)(1) - (-3)(-2) \end{pmatrix} = \begin{pmatrix} -17 \\ -18 \\ -4 \end{pmatrix}.$$

Lemma 1: Properties of the cross product.

Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^3$, with $\mathbf{z} = \mathbf{u} \times \mathbf{v}$, then:

(i) The vector $\mathbf{z}$ is orthogonal to both the vectors $\mathbf{u}$ and $\mathbf{v}$, that is,

$$\mathbf{z} \bullet \mathbf{u} = 0 \quad \text{and} \quad \mathbf{z} \bullet \mathbf{v} = 0.$$

(ii) We say that the cross product is **skew-symmetric**, that is,

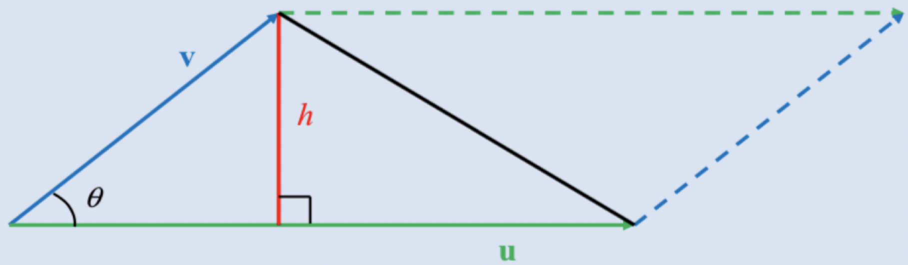
$$\mathbf{v} \times \mathbf{u} = -\,\mathbf{z} = -(\mathbf{u} \times \mathbf{v}).$$

(iii) If θ is the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$, then the length of $\mathbf{z}$ is

$$\|\mathbf{z}\| = \|\mathbf{u}\| \, \|\mathbf{v}\| \sin(\theta).$$

This is the area of the parallelogram formed by the two vectors $\mathbf{u}$ and $\mathbf{v}$.

Diagram 1



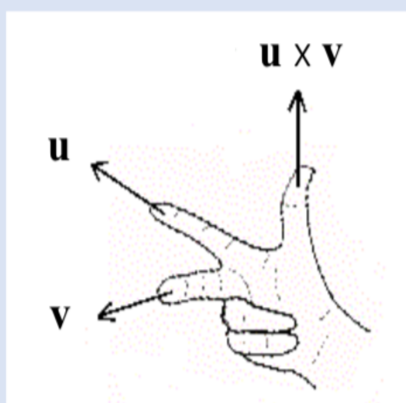
In the parallelogram and in the triangle formed by the vectors **u** and **v**, the height h of that triangle is given by: $h = \| \mathbf{v} \| \sin(\theta)$.

The area of the parallelogram is: $\| \mathbf{u} \| \, h = \| \mathbf{u} \| \, \| \mathbf{v} \| \sin(\theta) = \| \mathbf{z} \|$.

The area of the triangle is half of this last amount, i.e.: $\frac{1}{2} \| \mathbf{z} \|$.

(iv) The **right-hand rule:** this is used to find the direction of **z**. If the pointer finger of your right hand points in the direction of **u**, and the middle finger of your right hand points in the direction of **v**, then your thumb points in the direction of **z**.

Diagram 2



The **right-hand rule:** this is used to find the direction of the vector $\mathbf{z} = \mathbf{u} \times \mathbf{v}$.

If the pointer finger of your right hand points in the direction of **u**, and the middle finger of your hand right points in the direction of **v**, then your thumb points in the direction of **z**. (Thumb is pointing upwards)

Example 2

If $\mathbf{u} = \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}$, then from Example 1, $\mathbf{z} = \mathbf{u} \times \mathbf{v} = = \begin{pmatrix} -17 \\ -18 \\ -4 \end{pmatrix}$.

Consider

$$\mathbf{z} \bullet \mathbf{u} = \begin{pmatrix} -17 \\ -18 \\ -4 \end{pmatrix} \bullet \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} = -17(2) - 18(-3) - 4(5) = 0, \text{ and}$$

$$\mathbf{z} \bullet \mathbf{v} = \begin{pmatrix} -17 \\ -18 \\ -4 \end{pmatrix} \bullet \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} = -17(-2) - 18(1) - 4(4) = 0.$$

Thus we have verified property (i).

Notice that

$$\mathbf{v} \times \mathbf{u} = \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} \times \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} = \begin{pmatrix} (1)(5) - (4)(-3) \\ -((-2)(5) - (4)(2)) \\ (-2)(-3) - (1)(2) \end{pmatrix} = \begin{pmatrix} 17 \\ 18 \\ 4 \end{pmatrix} = -\mathbf{z}.$$

This verifies the skew-symmetry property (ii).

Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$, then we can evaluate θ from the dot product $\mathbf{u} \bullet \mathbf{v}$ (working to three decimal places throughout):

$$\theta = \arccos\left(\frac{\begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} \bullet \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}}{\| \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} \| \| \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} \|} \right) = \arccos\left(\frac{13}{\sqrt{798}} \right) = 1.093.$$

And so we can evaluate $\|\mathbf{u}\| \, \|\mathbf{v}\| \sin(\theta)$ (working to three decimal places throughout):

$$\| \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} \| \| \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix} \| \sin(\theta) = \sqrt{798}\,(0.888) = 25.080.$$

Whereas, we also have that $\mathbf{z} = \mathbf{u} \times \mathbf{v}$, and thus $\|\mathbf{z}\|$ is:

$$\|\mathbf{z}\| = \sqrt{17^2 + 18^2 + 4^2} = \sqrt{629} = 25.080.$$

We have just confirmed that the property (iii) holds.

Lemma 2: Linearity of the cross product

The cross product is linear in both arguments, that is, if $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in \mathbb{R}^3$, and $a \in \mathbb{R}$, then

$$\left.\begin{aligned} (\mathbf{x}+\mathbf{z}) \times \mathbf{y} &= (\mathbf{x} \times \mathbf{y}) + (\mathbf{z} \times \mathbf{y}) \\ (a\mathbf{x}) \times \mathbf{y} &= a\,(\mathbf{x} \times \mathbf{y}) \end{aligned}\right\}\text{: linearity in the first argument.}$$

$$\left.\begin{aligned} \mathbf{x} \times (\mathbf{z}+\mathbf{y}) &= (\mathbf{x} \times \mathbf{z}) + (\mathbf{x} \times \mathbf{y}) \\ \mathbf{x} \times (a\mathbf{y}) &= a(\mathbf{x} \times \mathbf{y}) \end{aligned}\right\}\text{: linearity in the second argument.}$$

Proof of Lemma 2: linearity in the second argument.

$$\mathbf{x} \times (\mathbf{z}+\mathbf{y}) = \begin{pmatrix} x_2(z_3+y_3) - x_3(z_2+y_2) \\ -[x_1(z_3+y_3) - x_3(z_1+y_1)] \\ x_1(z_2+y_2) - x_2(z_1+y_1) \end{pmatrix}$$

$$= \begin{pmatrix} x_2z_3 - x_3z_2 \\ -[x_1z_3 - x_3z_1] \\ x_1z_2 - x_2z_1 \end{pmatrix} + \begin{pmatrix} x_2y_3 - x_3y_2 \\ -[x_1y_3 - x_3y_1] \\ x_1y_2 - x_2y_1 \end{pmatrix} = (\mathbf{x} \times \mathbf{z}) + (\mathbf{x} \times \mathbf{y})$$

and

$$\mathbf{x} \times (a\mathbf{y}) = \begin{pmatrix} x_2(ay_3) - x_3(ay_2) \\ -[x_1(ay_3) - x_3(ay_1)] \\ x_1(ay_2) - x_2(ay_1) \end{pmatrix} = a \begin{pmatrix} x_2y_3 - x_3y_2 \\ -[x_1y_3 - x_3y_1] \\ x_1y_2 - x_2y_1 \end{pmatrix} = a(\mathbf{x} \times \mathbf{y}). \qquad \blacksquare$$

Topic 5

An Introduction to Linear Combinations and Span

Definition 1: Linear Combination

Let $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{F}^n$, and a and b be scalars in $\mathbb{F}$.

A **linear combination** of $\mathbf{v}$ and $\mathbf{w}$ means a vector of the form:

$$a\,\mathbf{v} + b\,\mathbf{w}.$$

Example 1

$$2\begin{pmatrix} 2 \\ 4 \\ 5 \end{pmatrix} - 5\begin{pmatrix} -3 \\ 7 \\ -9 \end{pmatrix},$$

is a linear combination of the vectors $\begin{pmatrix} 2 \\ 4 \\ 5 \end{pmatrix}$ and $\begin{pmatrix} -3 \\ 7 \\ -9 \end{pmatrix}$ in $\mathbb{R}^3$.

Example 2

$$(-2+3i)\begin{pmatrix} 2+i \\ 5-2i \end{pmatrix} + (4-2i)\begin{pmatrix} 7+8i \\ 5-4i \end{pmatrix},$$

is a linear combination of $\begin{pmatrix} 2+i \\ 5-2i \end{pmatrix}$ and $\begin{pmatrix} 7+8i \\ 5-4i \end{pmatrix}$ in $\mathbb{C}^2$.

Note that $0\mathbf{v} + 0\mathbf{w} = \mathbf{0}$, is a particular linear combination of $\mathbf{v}$ and $\mathbf{w}$.

The definition of a linear combination is extended in the obvious way to the case of more than 2 vectors in $\mathbb{F}^n$: that is,

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p, \quad \text{with } a_1, a_2, \ldots, a_p \in \mathbb{F},$$

is referred to as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{F}^n$.

Example 3

$$2\begin{pmatrix} -1 \\ 3 \\ -7 \\ 4 \\ 4 \end{pmatrix} - 5\begin{pmatrix} 2 \\ 5 \\ 7 \\ 9 \\ -6 \end{pmatrix} + 7\begin{pmatrix} 2 \\ -8 \\ 6 \\ -6 \\ 4 \end{pmatrix} - 4\begin{pmatrix} 5 \\ 7 \\ -4 \\ 2 \\ -2 \end{pmatrix} \left[= \begin{pmatrix} -18 \\ -103 \\ 9 \\ -87 \\ 74 \end{pmatrix} \right],$$

is a linear combination of the four vectors in $\mathbb{R}^5$:

$$(-1, 3, -7, 4, 4)^T, (2, 5, 7, 9, -6)^T, (2, -8, 6, -6, 4)^T, (5, 7, -4, 2, -2)^T.$$

Definition 2: Span

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ be vectors $\in \mathbb{F}^n$, then we define the **Span** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, written

$$\text{Span}(\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}),$$

to mean the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$: that is

$$\text{Span}(\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p : a_1, a_2, \ldots, a_p \in \mathbb{F}\}.$$

This is the set of all the vectors in $\mathbb{F}^n$ which you can construct from linear combinations of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. This is a very efficient way of producing a very large set of vectors from a relatively small set of vectors, and so we will make much use of this concept throughout the course.

Example 4: $\mathbb{F}$ could be $\mathbb{R}$ or $\mathbb{C}$ in this example.

Consider the vectors $\mathbf{v}_1 = (-1, 3, -7, 4, 4)^T$, $\mathbf{v}_2 = (2, 5, 7, 9, -6)^T$, $\mathbf{v}_3 = (2, -8, 6, -6, 4)^T$,

and $\mathbf{v}_4 = (5, 7, -4, 2, -2)^T$. Then $S = \text{Span}(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\})$, that is

$$S = \left\{ a\begin{pmatrix} -1 \\ 3 \\ -7 \\ 4 \\ 4 \end{pmatrix} + b\begin{pmatrix} 2 \\ 5 \\ 7 \\ 9 \\ -6 \end{pmatrix} + c\begin{pmatrix} 2 \\ -8 \\ 6 \\ -6 \\ 4 \end{pmatrix} + d\begin{pmatrix} 5 \\ 7 \\ -4 \\ 2 \\ -2 \end{pmatrix} : a, b, c, d \in \mathbb{F}. \right\}$$

Example 5: $\mathbb{F}$ must be $\mathbb{C}$ in this example since the vectors are in $\mathbb{C}^3$.

$$\operatorname{Span}\left(\left\{\begin{pmatrix} -1+2i \\ 3-4i \\ -7+9i \end{pmatrix}, \begin{pmatrix} 2+2i \\ 5 \\ 7-3i \end{pmatrix}, \begin{pmatrix} 2+i \\ -5+2i \\ 6i \end{pmatrix}\right\}\right)$$

$$= \left\{ w_1 \begin{pmatrix} -1+2i \\ 3-4i \\ -7+9i \end{pmatrix} + w_2 \begin{pmatrix} 2+2i \\ 5 \\ 7-3i \end{pmatrix} + w_3 \begin{pmatrix} 2+i \\ -5+2i \\ 6i \end{pmatrix} : w_1, w_2, w_3 \in \mathbb{C} \right\}.$$

Thinking ahead

One question that naturally arises: Is $\mathbf{v} \in \operatorname{Span}(\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\})$?

For example,

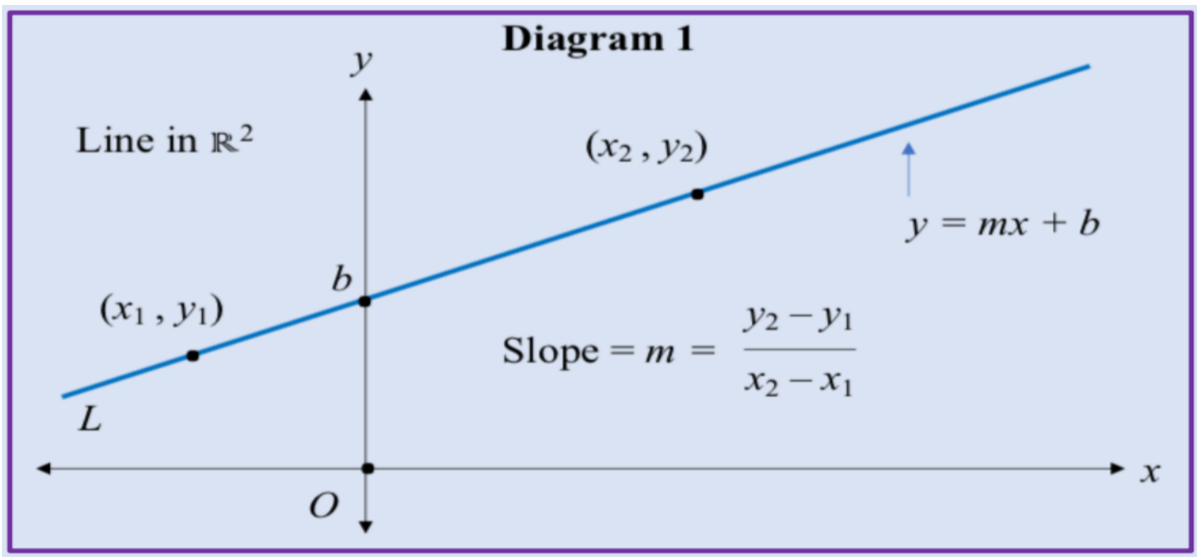
$$\text{Is } \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \in \operatorname{Span}\left(\left\{\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}\right)?$$

Topic 6A

Lines in $\mathbb{R}^n$

There is a nice way to write down the equation of a line in $\mathbb{R}^n$ making use of vectors. We will commence with some familiar material concerning lines in $\mathbb{R}^2$, and then move on to $\mathbb{R}^n$.

Lines in $\mathbb{R}^2$



There are 4 standard ways of providing the equation of a straight line in 2 dimensions.

(I) Slope (m) and y-intercept (b): with m and b fixed real numbers.

$$y = mx + b$$

(II) A Point, (x_1, y_1) and slope m.

$$y - y_1 = m(x - x_1)$$

(III) Two Points: (x_1, y_1) and (x_2, y_2)

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}.$$

Note that in these three forms, the variable "x" is preferred. We consider it as the input in the function, and it is used in order to obtain the output "y." That is, we have the form $y = y(x)$. However, there is really no reason to prefer "x", nor is there any reason to prefer "y".

(IV) Using a point, (x_1, y_1), a slope $\dfrac{q}{p}$ $(p \neq 0)$, and a parameter (often s or t),

$$x = x_1 + p\,t \quad \text{and} \quad y = y_1 + q\,t, \quad \text{with } t \in \mathbb{R}.$$

The real number t is the **parameter**.

In writing the equation of the line in case (IV), neither variable is preferred.
Note that there is freedom in choosing p and q, only their ratio is fixed.

Definition 1: Parametric equations of a line in $\mathbb{R}^2$

$$x = x_1 + p\,t \quad \text{and} \quad y = y_1 + q\,t, \quad \text{with } t \in \mathbb{R}.$$

We refer to the above expressions as the **parametric equations of the line in** $\mathbb{R}^2$ through the point (x_1, y_1) and slope $\dfrac{q}{p}$, where p, q are fixed real numbers and $p \neq 0$.

As t changes, you receive a different point on the line.
As t varies over all real numbers, you obtain all the points on the line.
Often t is, or can be thought of as, time.

For instance, if $t = 0$, then you have the point $x = x_1, y = y_1$,
if $t = 2$, then you have the point $x = x_1 + 2p, \quad y = y_1 + 2q$,
if $t = -5$, then you have the point $x = x_1 - 5p, \quad y = y_1 - 5q$.

Note that we have $p \neq 0$, so that $\dfrac{q}{p}$ *is* defined. If we do put $p = 0$ into the parametric equations of the line, then we get a line: it is vertical, with "infinite slope".

Definition 2: Vector equation of a line in $\mathbb{R}^2$

When we say the **vector equation of a line in** $\mathbb{R}^2$, we mean the expression:

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + t \begin{pmatrix} p \\ q \end{pmatrix}, \text{ for } t \in \mathbb{R}.$$

For any value of $t \in \mathbb{R}$, the expression $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + t \begin{pmatrix} p \\ q \end{pmatrix}$ will produce a vector in $\mathbb{R}^2$.

The terminal point associated to this vector, X, has coordinates $(x_1 + t\,p,\ y_1 + t\,q)$, $[= (x_1, y_1) + t\,(p, q)]$, and this is a point on the line.

If we introduce the vector $\mathbf{v}$ associated with the terminal point V, which has coordinates (x_1, y_1), and the vector $\mathbf{w}$ (note $\mathbf{w} \neq \mathbf{0}$) associated with the terminal point W, which has coordinates (p, q), then we can write the **vector equation of a line in** $\mathbb{R}^2$ as:

$$\mathbf{x} = \mathbf{v} + t\mathbf{w}\,, \text{ for } : t \in \mathbb{R}$$

We think of the "$\mathbf{v}$" part as taking you from the origin onto the line, i.e., to its terminal point V, and the "$t\,\mathbf{w}$" part as moving you along the line in the "$\mathbf{w}$" direction, by a scalar multiple of "t". We also say that $\mathbf{w}$ is tangential to the line, or that $\mathbf{w}$ is a tangent vector to the line.

The vector $\mathbf{w}$ is parallel to this line, however the terminal point W, associated with the vector $\mathbf{w}$, is not usually a point on the line. In fact, W is a point on the line iff the vector $\mathbf{v}$ is a multiple of the vector $\mathbf{w}$.

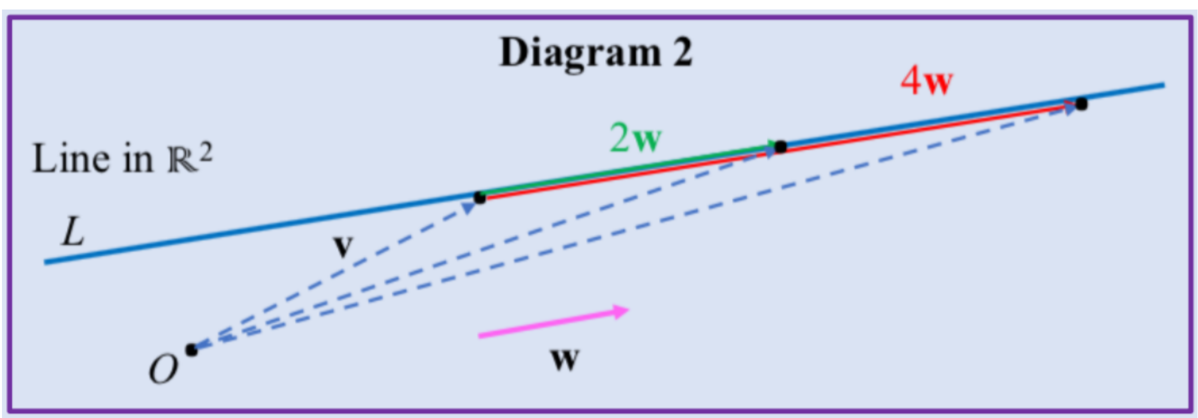
Definition 3: Line in $\mathbb{R}^2$

Let $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^2$ with $\mathbf{w} \neq \mathbf{0}$. We refer to the set of vectors

$$L = \{\mathbf{v} + t\,\mathbf{w} : t \in \mathbb{R}\}\,,$$

as the line in $\mathbb{R}^2$ passing through $\mathbf{v}$, with direction $\mathbf{w}$, or as just the line L, when $\mathbf{v}$ and $\mathbf{w}$ are given.

Really L is a set of vectors. Each one of these vectors has an associated terminal point. It is the set of terminal points which produces the set of points on the line.
The set of all points on the line is a subset of $\mathbb{R}^2$.



Lines in $\mathbb{R}^n$

Let $\mathbf{v}$ and $\mathbf{w}$ be two vectors in $\mathbb{R}^n, \mathbf{w} \neq \mathbf{0}$. We assume that the terminal point associated to $\mathbf{v}$ is the point V, with coordinates $(v_1, v_2, \ldots, v_n)$, and we assume that the terminal point associated to $\mathbf{w}$ is the point W, with coordinates $(w_1, w_2, \ldots, w_n)$.

Definition 4: Vector equation of a line in $\mathbb{R}^n$

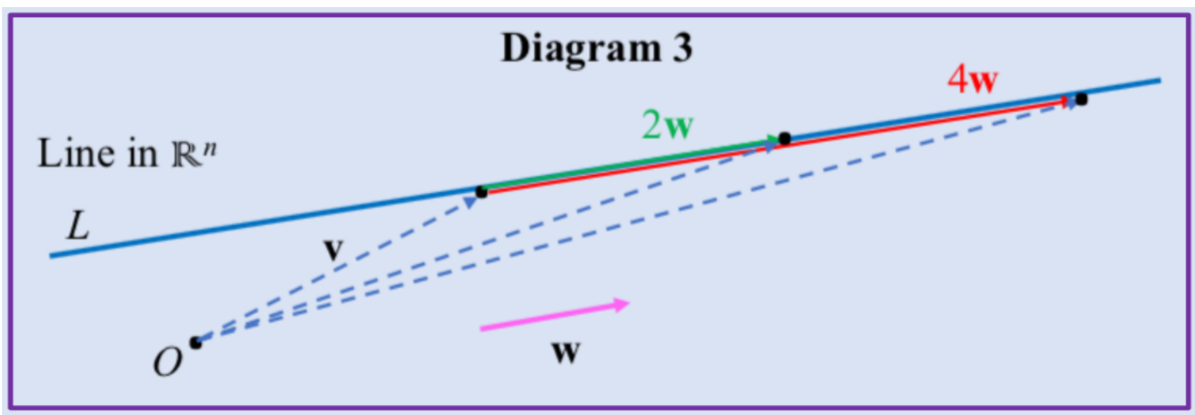
We say the **vector equation of a line in** $\mathbb{R}^n$, to mean the expression:

$$\mathbf{x} = \mathbf{v} + t\,\mathbf{w}\ , \text{ for } t \in \mathbb{R}, \text{ and } \mathbf{w} \neq \mathbf{0}.$$

For any value of $t \in \mathbb{R}$, $\mathbf{v} + t\,\mathbf{w}$ will produce a vector in $\mathbb{R}^n$. The terminal point X, associated to this vector, has coordinates $(v_1 + t\,w_1, v_2 + t\,w_2, \cdots, v_n + t\,w_n)$, $[\, = (v_1, v_2, \ldots, v_n) + t(w_1, w_2, \ldots, w_n)\ ]$, and this is a point on the line.

We think of the "$\mathbf{v}$" part as taking you from the origin onto the line, i.e., to its terminal point V, and the "$t\,\mathbf{w}$" part moves you along the line in the "$\mathbf{w}$" direction, by a scalar multiple of "t".

The vector $\mathbf{w}$ is parallel to this line, however the terminal point associated with the vector $\mathbf{w}$, W, is not usually a point on the line. In fact W is a point on the line iff the vector $\mathbf{v}$ is a multiple of the vector $\mathbf{w}$.



Notice that given particular vectors $\mathbf{v}$ and $\mathbf{w}$, then they will determine a unique line through the expression: $\mathbf{v} + t\,\mathbf{w}$. However, there are many other vectors which could also be used to produce that same line as a point different from V could have been given.

Definition 5: Parametric equations of a line in $\mathbb{R}^n$

Let consider the vector equation of the line in $\mathbb{R}^n$ given by :

$$\mathbf{x} = \mathbf{v} + t\,\mathbf{w}, \quad \text{for } t \in \mathbb{R}, \text{ and } \mathbf{w} \neq \mathbf{0}.$$

If we let the coordinates of the terminal point X in the Definition 4 to be $(x_1, x_2, \cdots, x_n)$,

then we obtain the **parametric equations of the line in** $\mathbb{R}^n$, namely

$$\begin{cases} x_1 = v_1 + t\,w_1 \\ x_2 = v_2 + t\,w_2 \\ \quad\vdots \\ x_n = v_n + t\,w_n \end{cases}.$$

Example 1

Give the vector equation and the parametric equations, of the line through the point V with coordinates $(2,-3,5)$ and pointing in the direction of the vector $\mathbf{w}=(-2,4,1)^T$.

Solution

The vector equation of the line is given by:

$$\mathbf{x}=\begin{pmatrix}2\\-3\\5\end{pmatrix}+t\begin{pmatrix}-2\\4\\1\end{pmatrix},\ t\in\mathbb{R}.$$

Note that the terminal point W (with coordinates $(-2,4,1)$) is not a point on this line.

The parametric equations of the line are:

$$\begin{cases} x=2-2\,t\\ y=-3+4\,t\\ z=5+t\end{cases},\quad \text{for } t\in\mathbb{R}.$$

Definition 6: Line in $\mathbb{R}^n$

Let $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ with $\mathbf{w}\neq\mathbf{0}$. We refer to the set of vectors

$$L=\{\mathbf{v}+t\,\mathbf{w}:t\in\mathbb{R}\},$$

as the line in $\mathbb{R}^n$ passing through $\mathbf{v}$, with direction $\mathbf{w}$, or as just the **line** L, when $\mathbf{v}$ and $\mathbf{w}$ are given.

Really L is a set of vectors. Each one of these vectors has an associated terminal point. It is the set of terminal points which produces the set of points on the line.

Example 2

$$\mathbf{x}=\begin{pmatrix}-2\\5\\7\end{pmatrix}+s\begin{pmatrix}6\\-12\\-3\end{pmatrix},\quad \text{for } s\in\mathbb{R},$$

is the vector equation of the same line as in Example 1.

We arrived at this conclusion:

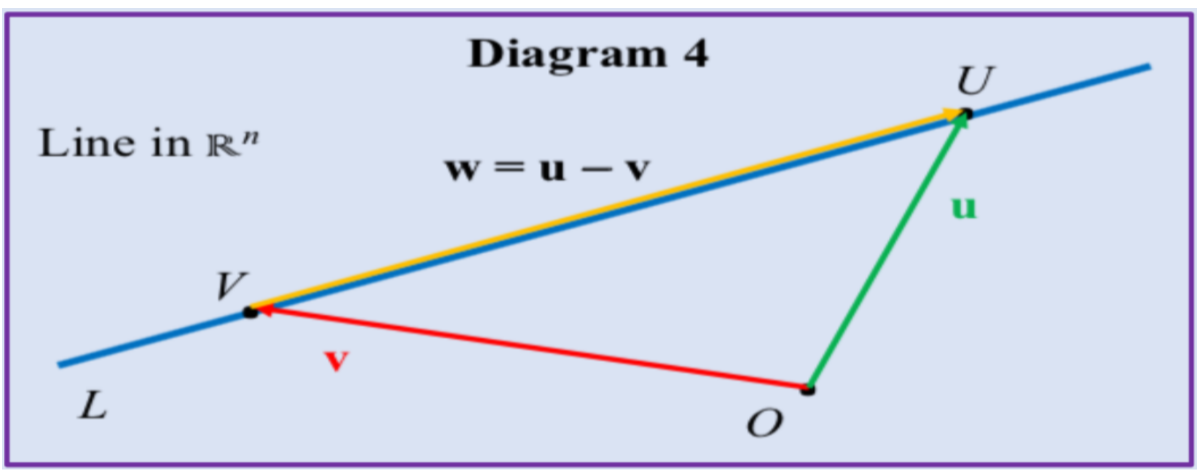
(1) by choosing $t=2$ in Example 1: we get the point $(-2,5,7)$, which is another point on the line, and

(2) by choosing the vector $-3\begin{pmatrix} -2 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ -12 \\ -3 \end{pmatrix}$, which has the same direction as the vector $\begin{pmatrix} -2 \\ 4 \\ 1 \end{pmatrix}$.

(The vector equation of the line in Example 1 is: $\mathbf{x} = \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} + t\begin{pmatrix} -2 \\ 4 \\ 1 \end{pmatrix}$, $t \in \mathbb{R}$.)

Suppose that you are now given two **distinct** points, V and U on a line in $\mathbb{R}^n$, with associated vectors $\mathbf{v}$ and $\mathbf{u}$, respectively. We can define $\mathbf{w} = \mathbf{u} - \mathbf{v}$, and this is a vector which gives the direction of the line. Thus we can still use Definition 4 for the vector equation of the line:

$$\mathbf{x} = \mathbf{v} + t\,(\mathbf{u} - \mathbf{v}), \quad t \in \mathbb{R}.$$



Example 3

What is the vector equation of the line through the two points $(2, -3, 5)$ and $(4, -2, 6)$?

Solution

The vector equation of the line through those points is :

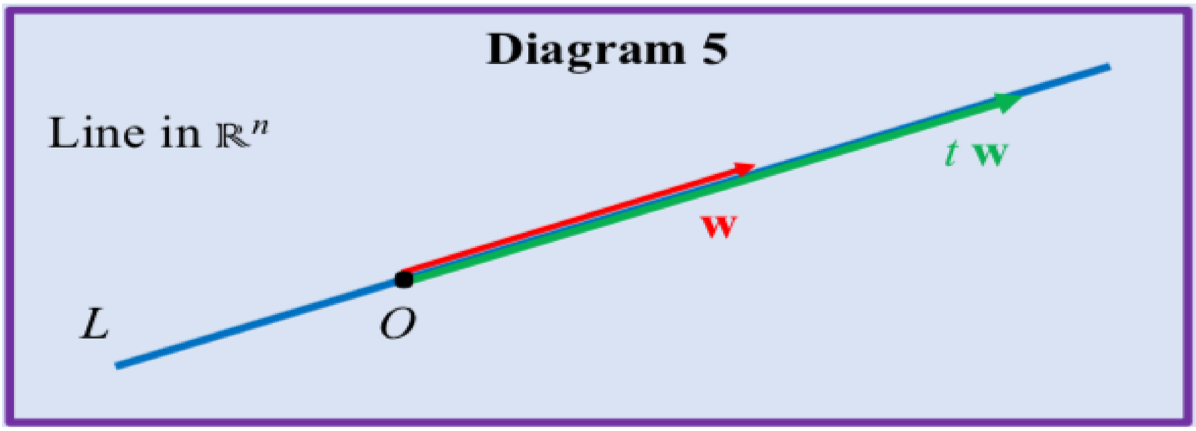
$$\mathbf{x} = \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} + t\left[\begin{pmatrix} 4 \\ -2 \\ 6 \end{pmatrix} - \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix}\right], \quad \text{for } t \in \mathbb{R},$$

that is,

$$\mathbf{x} = \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} + t\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \quad \text{for } t \in \mathbb{R}.$$

The Equation of a Line through the Origin in $\mathbb{R}^n$ using Span

If $\mathbf{w} \in \mathbb{R}^n$ with $\mathbf{w} \neq \mathbf{0}$, then $\mathrm{Span}(\{\mathbf{w}\}) = \{\mathbf{0} + t\,\mathbf{w} : t \in \mathbb{R}\}$, is a line. This is the straight line passing through O and in the $\mathbf{w}$ direction. This line itself is unique, however, as we know, there are many other ways in which we write down this line.



Note that we can only make use of the idea of Span to provide the equation of a line if the line passes through the origin.

Example 4

$$\mathrm{Span}\left(\left\{\begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}\right\}\right),$$

is a line passing through the origin in $\mathbb{R}^4$ pointing in the direction of the vector $(2, 4, 6, 8)^T$.

It is a shorthand for

$$L = \left\{ t \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Topic 6B

Planes in $\mathbb{R}^n$

The Equation of a Plane through the Origin in $\mathbb{R}^n$ using Span

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, and suppose we consider the set $P = \text{Span}(\{\mathbf{v}, \mathbf{w}\}) = \{s\,\mathbf{v} + t\,\mathbf{w} : t, s \in \mathbb{R}\}$. This is a subset of $\mathbb{R}^n$ containing, amongst other vectors, the vectors $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{0}$.
In $\mathbb{R}^3$, if $\mathbf{v} \neq \mathbf{0}, \mathbf{w} \neq \mathbf{0}$, $\mathbf{w} \neq m\mathbf{v}$, then P is the set of all vectors whose associated terminal points lie on a **plane** through the origin O. This plane contains the points V and W, which are the terminal points associated with the vectors $\mathbf{v}$ and $\mathbf{w}$ respectively. Also, the vectors $\mathbf{v}$ and $\mathbf{w}$ are tangent vectors to the plane. We extend this definition to $\mathbb{R}^n$.

Definition 7: Plane through the origin in $\mathbb{R}^n$

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}, \mathbf{w} \neq \mathbf{0}$, $\mathbf{w} \neq m\mathbf{v}$. We define a **plane through the origin in** $\mathbb{R}^n$, to mean

$$P = \text{Span}(\{\mathbf{v}, \mathbf{w}\}) = \{s\,\mathbf{v} + t\,\mathbf{w} : s, t \in \mathbb{R}\}\,.$$

Really P is a set of vectors. Each one of these vectors has an associated terminal point.

It is the set of terminal points which produces the set of points on the plane. If a point X, with associated vector $\mathbf{x}$, lies on the plane then we have $\mathbf{x} = s\,\mathbf{v} + t\,\mathbf{w}$, for some $s, t \in \mathbb{R}$; we give this expression a name.

Definition 8: Vector equation of a plane in $\mathbb{R}^n$

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}, \mathbf{w} \neq \mathbf{0}$, $\mathbf{w} \neq m\mathbf{v}$. We refer to the expression:

$$\mathbf{x} = s\,\mathbf{v} + t\,\mathbf{w},$$

as the **vector equation of a plane through the origin** in $\mathbb{R}^n$.

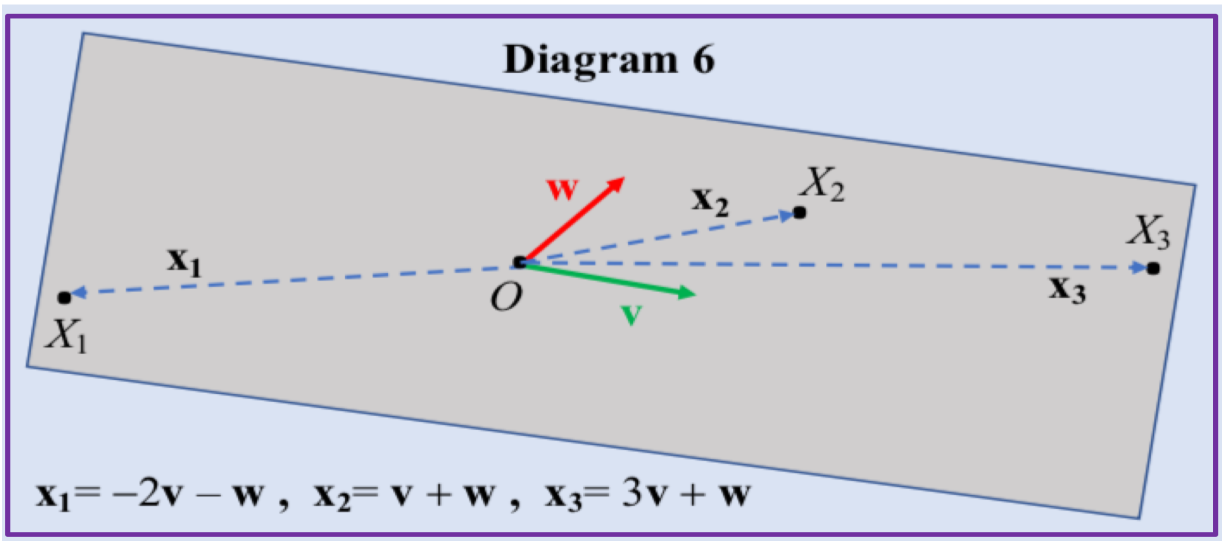
We can use, "the plane (through the origin)," to refer to either $s\,\mathbf{v} + t\,\mathbf{w}$ or to $\text{Span}(\{\mathbf{v}, \mathbf{w}\})$.

The parameters s and t "move" you in the plane in the $\mathbf{v}$ and $\mathbf{w}$ directions respectively. The plane contains the points V and W, which are the terminal points associated with the vectors $\mathbf{v}$ and $\mathbf{w}$ respectively. Also, the vectors $\mathbf{v}$ and $\mathbf{w}$ are tangent vectors to the plane.

Notice, also that if any one of the conditions, $\mathbf{v} \neq \mathbf{0}$, or $\mathbf{w} \neq \mathbf{0}$ or $\mathbf{w} \neq m\mathbf{v}$ is violated, then we do not have a plane. If either $\mathbf{v}$ or $\mathbf{w}$ (but not both) is $\mathbf{0}$, or one vector is a

multiple of the other vector, then P is a line.
Clearly, if both $\mathbf{v} = \mathbf{0}$ and $\mathbf{w} = \mathbf{0}$, then P is just a point - the zero vector.

Notice that given $\mathbf{v}$ and $\mathbf{w}$, then the expression $s\,\mathbf{v} + t\,\mathbf{w}$ produces a unique plane, however, there are many other vectors which can be used to produce that same plane.



Note that we can only make use of the idea of Span to provide the equation of a plane if the plane passes through the origin.

Example 5

Give the vector equation of a plane in $\mathbb{R}^3$ which passes through the origin, and has the two vectors $\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix}$ tangential to it.

Solution

The vector equation of this plane is:

$$\mathbf{x} = s\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} + t\begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix}, \quad \text{for } s, t \in \mathbb{R}.$$

Notice that the points $(2, 4, 6)$ and $(-1, 2, -3)$ are on this plane.

Example 6

Give the vector equation of a plane in $\mathbb{R}^5$ which passes through the origin, and has the two vectors $(5,4,3,2,1)^T$ and $(-5,4,-3,2,-1)^T$ tangential to it.

Solution

The vector equation of this plane is:

$$\mathbf{x} = s\begin{pmatrix} 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{pmatrix} + t\begin{pmatrix} -5 \\ 4 \\ -3 \\ 2 \\ -1 \end{pmatrix}, \quad \text{for } s,t \in \mathbb{R}$$

Notice that the points $(5,4,3,2,1)$ and $(-5,4,-3,2,-1)$ in $\mathbb{R}^5$, are points on the plane.

The Vector Equation of a Plane in $\mathbb{R}^n$

Definition 9: Plane in $\mathbb{R}^n$

Let $\mathbf{p}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}, \mathbf{w} \neq \mathbf{0}$, $\mathbf{w} \neq m\,\mathbf{v}$. We define a **plane** in $\mathbb{R}^n$ to mean

$$P = \{\mathbf{p} + s\,\mathbf{v} + t\,\mathbf{w} : s,t \in \mathbb{R}\}.$$

Really P is a set of vectors. Each one of these vectors has an associated terminal point.

It is the set of terminal points which produces the set of points on the plane.

If a point X with associated vector $\mathbf{x}$ lies on the plane, we then have:

$$\mathbf{x} = \mathbf{p} + s\,\mathbf{v} + t\,\mathbf{w}, \quad \text{for some } s,t \in \mathbb{R}.$$

We give this expression a name.

Definition 10: Vector equation of a plane

Let $\mathbf{p}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}, \mathbf{w} \neq \mathbf{0}$, $\mathbf{w} \neq m\,\mathbf{v}$, we refer to the expression:

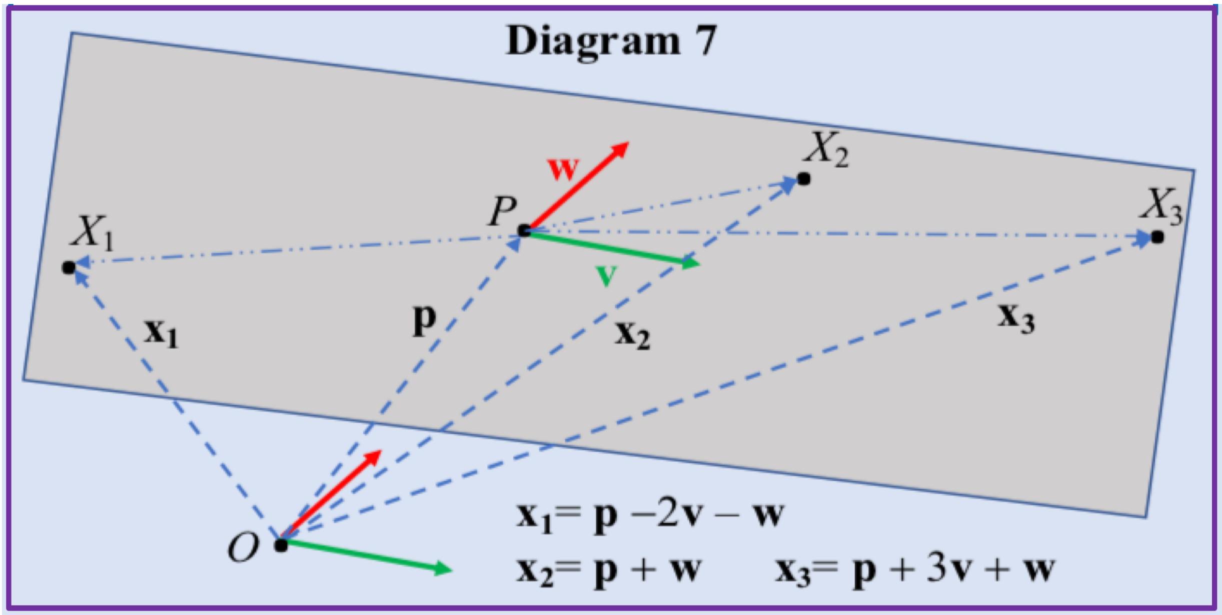
$$\mathbf{x} = \mathbf{p} + s\,\mathbf{v} + t\,\mathbf{w},$$

as the **vector equation of a plane**.

We can use "the plane" to refer to either $\mathbf{p} + s\,\mathbf{v} + t\,\mathbf{w}$, or $\{\mathbf{p} + s\,\mathbf{v} + t\,\mathbf{w} : s,t \in \mathbb{R}\}$.

Note that this is **not** simply a "Span" of something.

The parameters s and t "move" you in the plane in the $\mathbf{v}$ and $\mathbf{w}$ directions respectively. The vectors $\mathbf{v}$ and $\mathbf{w}$ are tangent vectors to the plane. The plane does not usually contain the points V and W, which are the terminal points associated with the vectors $\mathbf{v}$ and $\mathbf{w}$, respectively. In fact V and W are points on the plane iff $\mathbf{p} \in \text{Span}\left(\{\mathbf{v}, \mathbf{w}\}\right)$.



Notice that given $\mathbf{p}, \mathbf{v}$ and $\mathbf{w}$, then the expression $\mathbf{p} + s\,\mathbf{v} + t\,\mathbf{w}$ produces a unique plane, however, there are many other vectors which can be used to produce that same plane.

Example 7

Give the vector equation of a plane in $\mathbb{R}^3$ which passes through the point $(1, -4, 3)$,

and has the two vectors $\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix}$ tangential to it.

Solution

The vector equation of this plane is:

$$\mathbf{x} = \begin{pmatrix} 1 \\ -4 \\ 3 \end{pmatrix} + s\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} + t\begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix}, \quad \text{for some } s, t \in \mathbb{R}.$$

Notice that the points $(2, 4, 6)$ and $(-1, 2, -3)$ are NOT on this plane, however, the vectors $(2, 4, 6)^T$ and $(-1, 2, -3)^T$ are tangential to this plane.

Example 8

Give the vector equation of a plane in $\mathbb{R}^5$ which passes through the point $(1,-4,5,-2,7)$, has the two vectors, $(2,4,6,8,10)^T$ and $(-1,2,-3,4,-5)^T$, tangential to it.

Solution

$$\mathbf{x} = \begin{pmatrix} 1 \\ -4 \\ 5 \\ -2 \\ 7 \end{pmatrix} + s\begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \\ 10 \end{pmatrix} + t\begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \\ -5 \end{pmatrix}, \quad \text{for some } s,t \in \mathbb{R}.$$

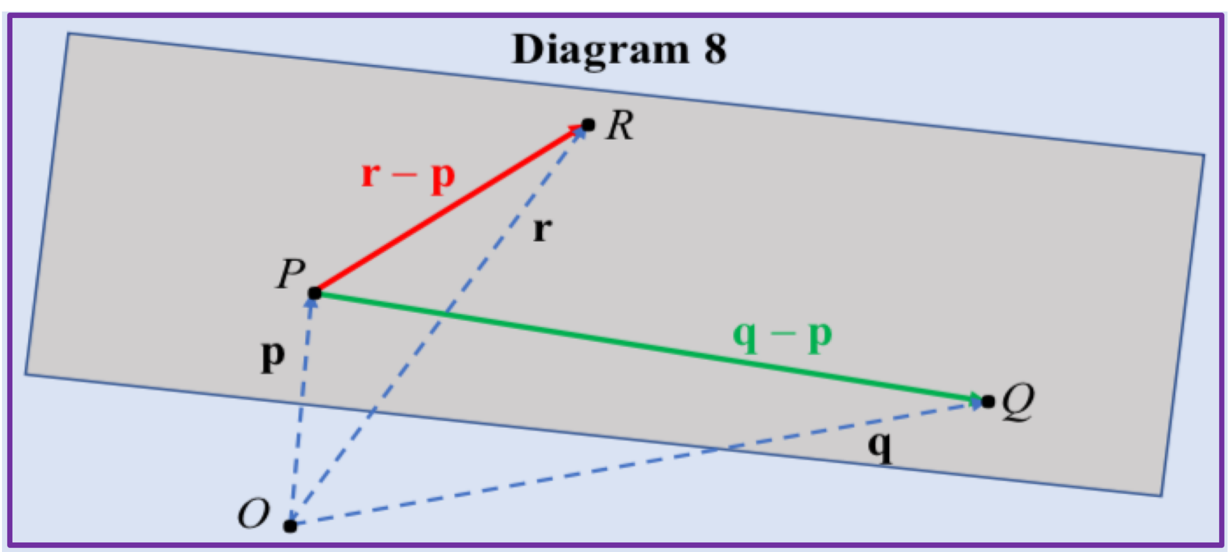
Notice that the points $(2,4,6,8,10)$ and $(-1,2,-3,4,-5)$ are NOT on the plane, however, the vectors $(2,4,6,8,10)^T$ and $(-1,2,-3,4,-5)^T$ are tangential to the plane. Simplifying the vector $(2,4,6,8,10)^T$, we could write the following equivalent solution:

$$\mathbf{x} = \begin{pmatrix} 1 \\ -4 \\ 5 \\ -2 \\ 7 \end{pmatrix} + s\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} + t\begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \\ -5 \end{pmatrix}, \quad \text{for some } s,t \in \mathbb{R}.$$

The Equation of a Plane in $\mathbb{R}^n$ using three Non-Colinear Points.

Let P, Q and R be three non-colinear points in $\mathbb{R}^n$, with associated vectors $\mathbf{p}$, $\mathbf{q}$, and $\mathbf{r}$, respectively. Suppose that we want to obtain the equation of the unique plane containing these three points. The two vectors $\mathbf{v} = \mathbf{q} - \mathbf{p}$ and $\mathbf{w} = \mathbf{r} - \mathbf{p}$ are tangential to the plane, and thus, we can express the equation of the plane as:

$$\Pi = \{\mathbf{p} + s(\mathbf{q} - \mathbf{p}) + t(\mathbf{r} - \mathbf{p}) : s,t \in \mathbb{R}\}.$$



Example 9

Find the vector equation of a plane in $\mathbb{R}^4$ which contains the following points:

$$(2,-4,6,-8),\ (1,3,-2,-4),\ \text{and } (9,7,5,3).$$

Solution

We first evaluate the difference vectors:

$$(1,3,-2,-4)^T - (2,-4,6,-8)^T = (-1,7,-8,4)^T$$

and

$$(9,7,5,3)^T - (2,-4,6,-8)^T = (7,11,-1,11)^T.$$

We then obtain the vector equation of this plane :

$$\mathbf{x} = \begin{pmatrix} 2 \\ -4 \\ 6 \\ -8 \end{pmatrix} + s \begin{pmatrix} -1 \\ 7 \\ -8 \\ 4 \end{pmatrix} + t \begin{pmatrix} 7 \\ 11 \\ -1 \\ 11 \end{pmatrix}, \quad \text{for } s,t \in \mathbb{R}.$$

Scalar Equation of a Plane in $\mathbb{R}^3$

The above expression for the equation of a plane do hold, in particular, in $\mathbb{R}^3$.
The following expression makes use of the cross product and thus holds **only** in $\mathbb{R}^3$.

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ with $\mathbf{v} \neq \mathbf{0}, \mathbf{w} \neq \mathbf{0}$, $\mathbf{w} \neq m\mathbf{v}$. We know that their cross product, $\mathbf{n} = \mathbf{v} \times \mathbf{w}$ is orthogonal to any plane to which $\mathbf{v}$ and $\mathbf{w}$ are tangential. The vector $\mathbf{n}$ is referred to as a **normal** (or as a normal vector) to the plane.

Suppose that P is a point on the plane, with associated vector $\mathbf{p}$, and X is any point in the plane, with associated vector $\mathbf{x}$, then the vector $\mathbf{x} - \mathbf{p}$ is tangential to the plane and thus is orthogonal to $\mathbf{n}$: that is, $\mathbf{n} \bullet (\mathbf{x} - \mathbf{p}) = (\mathbf{v} \times \mathbf{w}) \bullet (\mathbf{x} - \mathbf{p}) = 0$.

Definition 11: Scalar equation of a plane

Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ with $\mathbf{v} \neq \mathbf{0}, \mathbf{w} \neq \mathbf{0}$, $\mathbf{w} \neq m\mathbf{v}$, be vectors in $\mathbb{R}^3$.
Let $\mathbf{p}$ be a vector in $\mathbb{R}^3$. We refer to the expression

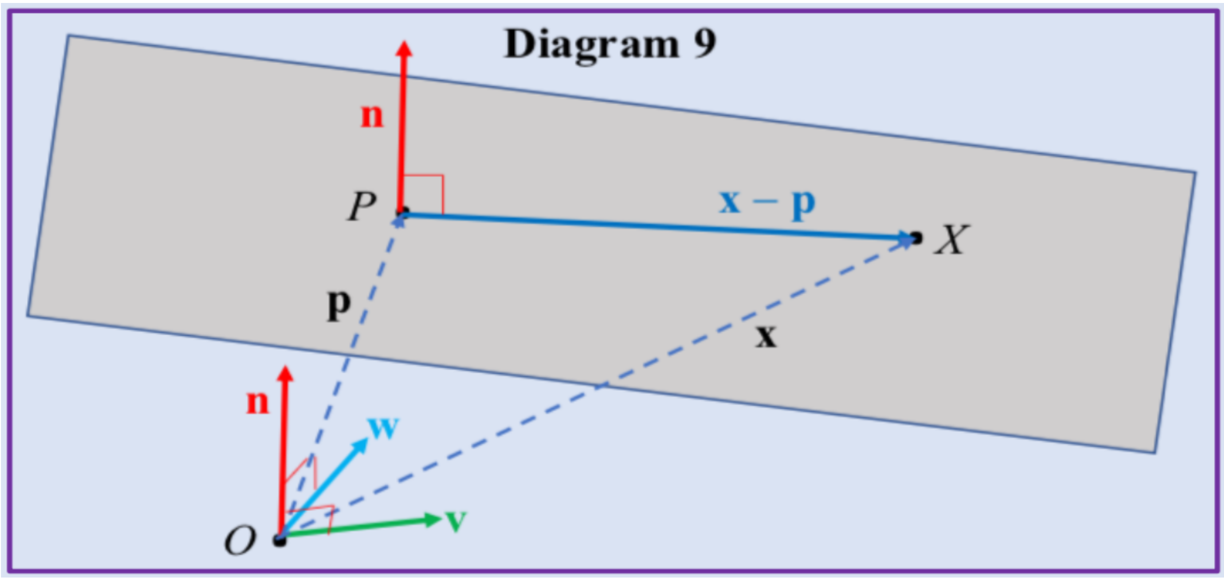
$$\mathbf{n} \bullet (\mathbf{x} - \mathbf{p}) = (\mathbf{v} \times \mathbf{w}) \bullet (\mathbf{x} - \mathbf{p}) = 0$$

as the **scalar equation of the plane in** $\mathbb{R}^3$, passing through the point P associated to the vector $\mathbf{p}$, and having the vectors $\mathbf{v}$ and $\mathbf{w}$ tangential to it.

This expression is the **scalar equation of a plane**, as the dot product gives us a scalar.

Note that the plane goes through the origin O iff the vector $\mathbf{0}$ satisfies this equation, that is, iff $(\mathbf{v} \times \mathbf{w}) \bullet (\mathbf{0} - \mathbf{p}) = 0$, i.e. iff $\mathbf{p} = a\,\mathbf{v} + b\,\mathbf{w}$, for some $a, b \in \mathbb{R}$.

This makes sense geometrically: the plane goes through the origin iff both V and W lie on the plane.



Example 10

Give the scalar equation of the plane in $\mathbb{R}^3$ which passes through the origin, and has the two vectors $(2, 4, 7)^T$and $(-1, 2, -3)^T$ tangential to it.

Solution

Let $\mathbf{n}$ be a normal to the plane and $\mathbf{x} = (x, y, z)^T$ be any vector on the plane.

We then have:

$$\mathbf{n} = \begin{pmatrix} 2 \\ 4 \\ 7 \end{pmatrix} \times \begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix} = \begin{pmatrix} -26 \\ -1 \\ 8 \end{pmatrix}$$

The scalar equation of the plane (going through the origin) is thus given by

$$\mathbf{x} \bullet \begin{pmatrix} -26 \\ -1 \\ 8 \end{pmatrix} = 0,$$

which simplifies to

$$-26x - y + 8z = 0.$$

Note that the two points $(2, 4, 7)$ and $(-1, 2, -3)$ lie on this plane.

Example 11

Give the scalar equation of the plane in $\mathbb{R}^3$ which passes through the point $(1,-4,3)$, and has the two vectors $(2,4,7)^T$ and $(-1,2,-3)^T$ tangential to it.

Solution

Let $\mathbf{n}$ be a normal to the plane and $\mathbf{x} = (x,y,z)^T$ be any vector on the plane.

$$\mathbf{n} = \begin{pmatrix} 2 \\ 4 \\ 7 \end{pmatrix} \times \begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix} = \begin{pmatrix} -26 \\ -1 \\ 8 \end{pmatrix}.$$

The scalar equation of the plane (not going through the origin) is then given by

$$\left[\mathbf{x} - \begin{pmatrix} 1 \\ -4 \\ 3 \end{pmatrix}\right] \bullet \begin{pmatrix} -26 \\ -1 \\ 8 \end{pmatrix} = 0 = -26(x-1) - (y+4) + 8(z-3),$$

which simplifies to

$$-26\,x - y + 8\,z = 2.$$

Note that the two points $(2,4,7)$ and $(-1,2,-3)$ do not lie on this plane.

Topic 7A

Systems of Linear Equations

We begin with some simple examples in which linear systems of equations naturally arise.

Example 1

10 apples cost \$3.50, how much does one apple cost?

Solution

Let a denote the price of one apple, in cents. Then, converting dollars to cents, we can write the information in mathematical notation as:

$$10\,a = 350$$

$$\Rightarrow a = \frac{350}{10} = 35.$$

We conclude that the price of an apple is 35 cents.

Example 2

At the market, 6 apples and 4 peaches cost \$4, and 4 apples and 6 peaches cost \$5. How much does an apple cost and a peach cost?

Solution

Let a denote the price of an apple in cents, and p denote the price of a peach in cents. We write down the provided information in mathematical form,

$$\begin{cases} 6\,a + 4\,p = 400 & (e_1) \\ 4\,a + 6\,p = 500 & (e_2) \end{cases}$$

Notice that the large curly bracket in the system of equations stands for "and" as **both** equations, labelled (e_1) and (e_2), must be satisfied simultaneously.

Consider now $6\,(e_1) - 4\,(e_2)$ to get:

$$([6(6) - 4(4)]\,a + [6(4) - 4(6)]\,p = 6(400) - 4(500), \quad \text{that is}$$

$$20\,a = 400 \quad \Rightarrow \quad a = \frac{400}{20} = 20.$$

Substituting this value for a in equation (e_1) gives $p = \dfrac{(400-120)}{4} = 70.$

You could use (e_2) and get the same result. A good question to ask is, why?

We conclude that the price of an apple is 20 cents and the price of a peach is 70 cents.

Comment 1
The crucial step in solving this problem is that of taking the combination $6\,(e_1) - 4\,(e_2)$. This step is determined by looking at the coefficients of p in the two equations (e_1) and (e_2), which are 4 and 6, respectively. The combination of $6\,(e_1) - 4\,(e_2)$ produces a new equation which has **no** p terms in it. We then solve that new equation for the other variable, a. Once we have a value for a, then we return to one of the original equations (i.e. (e_1) or (e_2)), insert this value for a, and then use the equation to determine the value of the other variable p.

Note that we could have alternatively considered the combination $4\,(e_1) - 6\,(e_2)$, to eliminate first the variable a, solve for p, and then replace its value in either equation, (e_1) or (e_2), in order to find the value of the other variable, a.

Comment 2
We should never give an incorrect answer when solving a system of equations, because we can always check that our solution does indeed satisfy the original system. In this case, if we put the values of 20 for a and 70 for p into the two equations we get:

$$6(20) + 4(70) = 400 \text{ - this is correct}$$
$$4(20) + 6(70) = 500 \text{ - this is correct}$$

In each case, once we have tidied the expressions up, we get the value that originally appeared on the right hand side (RHS) of the equation.

Example 3

How to find the equation of the line that passes through the points $(2,3)$ and $(-3,4)$.

Solution

We must find the constants m and b, in the equation $y = mx + b$, so that:

$$3 = 2m + b, \text{ and } 4 = -3m + b.$$

We thus have the system of two equations in the two unknowns, m and b:

$$\begin{cases} 2\,m + b = 3 & (e_1) \\ -3\,m + b = 4 & (e_2) \end{cases}$$

We will not solve this system now.

Example 4

How to determine if the vector $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ lies in **Span** $\left(\left\{\begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}\right)$.

Solution

We must then find the scalars $a, b \in \mathbb{R}$, such that

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = a \begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

or equivalently

$$\begin{cases} 1 = (-1)\,a + (1)\,b & (e_1) \\ 2 = (2)\,a + (1)\,b & (e_2) \\ 3 = (-3)\,a + (1)\,b & (e_3) \end{cases}$$

In this problem, we are led to a linear system of 3 equations in 2 unknowns, that is:

$$\begin{cases} 1 = -a + b & (e_1) \\ 2 = 2\,a + b & (e_2) \\ 3 = -3\,a + b & (e_3) \end{cases}$$

We will not solve this problem now.

Definition 1 : Linear equation

An equation in a single variable, x, is called **linear** to mean that the unknown, x, in the equation can **only** occur linearly, that is, to the power of 0 or 1.

If x is an unknown then there will be no terms of $x^2, x^3, \frac{1}{x}, \ln(x), \sin(x), e^x$, etc.
There can be terms that *do not* involve the variable x at all, and there can be terms in x^1.

An equation in the unknowns $x_1, \cdots, x_n$, is called **linear in each of the unknowns** $x_1, \cdots, x_n$, (or just **linear**), to mean that the equation is linear in each unknown x_i, $(i = 1, \cdots, n)$, and that there are no products of these unknowns (e.g. $x_1\,x_2$).

Definition 2: Linear system of m equations in n unknowns

The general form of a linear system of m equations in n unknowns is:

$$(*) \quad \begin{cases} a_{11}\,x_1 + a_{12}\,x_2 + \cdots + a_{1n}\,x_n = b_1 & (e_1) \\ a_{21}\,x_1 + a_{22}\,x_2 + \cdots + a_{2n}\,x_n = b_2 & (e_2) \\ \vdots & \\ a_{m1}\,x_1 + a_{m2}\,x_2 + \cdots + a_{mn}\,x_n = b_m & (e_m) \end{cases}$$

We will refer to this system of equations as $(*)$ for short.

Definition 3: Coefficients and unknowns

The scalars a_{ij} in $\mathbb{F}$ are called the **coefficients** in the system, these are usually known.

The variables $x_1, x_2, \ldots, x_n$ in $\mathbb{F}$ are referred to as the **unknowns.**

The variables $b_1, b_2, \ldots, b_m$ in $\mathbb{F}$ are often referred to collectively as the right-hand side (RHS) (of the equation).

Definition 4: Solve and solution

We say that the scalars $y_1, y_2, \ldots, y_n$ in $\mathbb{F}$, **solve** the system $(*)$ to mean that if we set $x_1 = y_1, x_2 = y_2, \ldots, x_n = y_n$ in the system, then we get an identity. That is, each of the equations is identically satisfied.

$$(*)\quad \left\{ \begin{array}{cc} a_{11}\, y_1 + a_{12}\, y_2 + \cdots + a_{1n}\, y_n = b_1 & (e_1) \\ a_{21}\, y_1 + a_{22}\, y_2 + \cdots + a_{2n}\, y_n = b_2 & (e_2) \\ \vdots & \\ a_{m1}\, y_1 + a_{m2}\, y_2 + \cdots + a_{mn}\, y_n = b_m & (e_m) \end{array} \right.$$

We also say that (the vector) $\mathbf{x} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ is a **solution** to $(*)$.

Otherwise we will say that $\mathbf{x}$ is not a solution to $(*)$.

Note that if just one equation in the system $(*)$ is NOT satisfied, then $\mathbf{x} = (y_1, y_2, \ldots, y_n)^T$ is not a solution.

Example 5

Consider the following system of linear equations:

$$\left\{ \begin{array}{rl} x_1 - 2x_2 + 3x_3 &= 1 \\ 2x_1 - 4x_2 + 6x_3 &= 2 \\ 3x_1 - x_2 + x_3 &= 3, \end{array} \right.$$

then, $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is a solution, as

$$\left\{ \begin{array}{l} 1(1) - 2(0) + 3(0) = 1 \\ 2(1) - 4(0) + 6(0) = 2 \\ 3(1) - 1(0) + 1(0) = 3. \end{array} \right.$$

However, $\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$ is not a solution, as

$$\begin{aligned} &1(-2) - 2(0) + 3(1) = 1, \text{ and} \\ &2(-2) - 4(0) + 6(1) = 2, \textbf{ but} \\ &3(-2) - 1(0) + 1(1) = -5 \neq 3. \end{aligned}$$

Definition 5: Solution set

We will want to find the set of all solutions to $(*)$, this is called the **solution set** to $(*)$, and we will denote it by S.

We will often have some work to do in order to obtain S.

Example 6

Solve $2\,x = 4$.

Solution

The solution $x = 2$ is easily found, and thus the solution set S is $\{2\}$.

We check: $2(2) = 4$, which is the RHS of the equation, so we are correct.

Example 7

Solve $2\,x + 4\,y = 6$.

Solution

This is a little harder. We only have one equation for the two unknowns, x and y.
We "feel" that we are missing some information. We cannot obtain just one solution, and there will be some uncertainty in our solution. The solution will reflect this uncertainty, and contain a parameter.

We could let $y = t$, say, with $t \in \mathbb{R}$, that is, we are thinking that y can be anything, and we are using t to refer to that anything. We then solve the single equation for x :

$$2\,x + 4\,t = 6 \Rightarrow x = 3 - 2\,t, \text{ and the solution set is } \left\{ \begin{pmatrix} 3 - 2\,t \\ t \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Let us check that we have indeed solved correctly the equation $2\,x + 4\,y = 6$.
If we let $y = t$, and $x = 3 - 2\,t$ in the LHS of the equation, then we have:

$$2(3-2t)+4t=6-4t+4t=6,$$

which is the RHS of the equation; so we have a solution.

Note that at this time, we do not know if we are missing any solution(s); we will find out later that we are not.
Notice that you could just let $y=y$, (instead of $y=t$) and then get $x=3-2y$. However, this makes y look preferred to x to some extent, and there is no reason for this to be the case. This is why it is traditional to introduce a parameter t, either letting $y=t$ and solve for x, or letting $x=t$ and solve for y.

Example 8

Solve the system of linear equations:

$$\begin{cases} 2x+3y=4 \\ 5x-7y=18 \end{cases}$$

This is much more difficult.
We will discuss techniques for solving such problems in the next few weeks.

Theorem 1: The solution set to a linear system of equations.

The solution set to system $(*)$ is either:

(a) empty - there are no solutions

(b) contains exactly one element - there is a unique solution, or

(c) contains an infinite number of elements - there are parameters in the solution set.

Definition 6: Inconsistent and consistent systems

If the solution set is empty, then we say that the system is **inconsistent.**

If the solution set has a unique solution or infinitely many solutions, then we say that the system is **consistent.**

Example 9

a) Solve the system: $\begin{cases} x=1, \\ x=2, \end{cases}$

Clearly there is no solution and $S=\emptyset$.

b) Solve the system: $\begin{cases} x + y = 2, \\ x - y = 4, \end{cases}$

In this case, the unique solution is: $x = 3,\ y = -1$. And $S = \left\{ \begin{pmatrix} 3 \\ -1 \end{pmatrix} \right\}$.

c) Solve the system: $x + y = 1$.

Let $y = t$, where $t \in \mathbb{R}$, then $x = 1 - t$.

We obtain thus infinitely many solutions and $S = \left\{ \begin{pmatrix} 1-t \\ t \end{pmatrix} : t \in \mathbb{R} \right\}$.

Note if we let instead $x = t$, where $t \in \mathbb{R}$ and $y = 1 - t$, then $S = \left\{ \begin{pmatrix} t \\ 1-t \end{pmatrix} : t \in \mathbb{R} \right\}$.

Note that it is usually not obvious by looking at the system whether it will be inconsistent or not, and if consistent, whether the system will have a unique solution or infinitely many solutions. Expect, thus, that some calculations will be required.

So when solving a system of equations $(*)$, what can we do to obtain its solution set?
The key idea is that we manipulate the original system $(*)$ into another system, which we will refer to as $(**)$, and which has the same solution set as $(*)$. This will be useful if the solution set of $(**)$ is easier to obtain than the solution set of $(*)$.
This is what we will investigate more closely next.

Definition 7: Equivalent systems

We say that two linear systems are **equivalent** to mean that they have the same solution set.

Thus given a system of linear equations $(*)$ we will manipulate it into equivalent systems, and we will stop once we have obtained an equivalent system whose solution set is easily obtained.

What manipulations are then allowed, and which ones are not?

Example 10:

(a) Consider the system $(*)$ $\begin{cases} x + y = 1 & (e_1) \\ 2x + 3y = 6 & (e_2) \end{cases}$

If we multiply the first equation by zero, we get a new system $(**)$:

$$(**) \quad \begin{cases} 0 = 0 & (e_3) \\ 2x + 3y = 6 & (e_2) \end{cases}$$

.

Clearly $(**)$ is not equivalent to $(*)$.
We have lost information by **multiplying** (e_1) **by zero** to produce the new equation (e_3).

(b) Consider now the system $(*)$ $\begin{cases} x + y + z = 1 & (e_1) \\ 2x + 3y - 2z = 6 & (e_2) \end{cases}$

Let us now **add an equation to** $(*)$ to get a new system $(**)$:

$$(**) \quad \begin{cases} x + y + z = 1 & (e_1) \\ 2x + 3y - 2z = 6 & (e_2) \\ x + 2y - 3z = 5 & (e_3) \end{cases}$$

We suspect that this new system $(**)$ is not equivalent to the old one $(*)$, since we have added an extra restriction on the variables, namely the new equation (e_3).

We must be careful that we neither lose information nor add information.

Topic 7B

Systems of Linear Equations

(Continuation of Topic 7A)

There are three basic operations that we perform on our system of linear equations which will produce an equivalent system. These basic operations are called Elementary Operations.

Definition 8: Elementary Operations

Let us consider a system $(*)$ of m linear equations in n variables which are labelled and ordered from (e_1), for the first equation, to (e_m), for the last equation.
In order to obtain an equivalent system, one of the following three operations can be performed one at a time on the system of equations.
These operations are known as **Elementary Operations** of the indicated type.

Type I Elementary Operation: **interchange two equations**

$e_i \leftrightarrow e_j$ Interchange equations i and j.

Type II Elementary Operation: **multiply one equation by a non-zero constant**

$e_i \to k\, e_i,\ k \in \mathbb{F} \setminus \{0\}$ Replace equation i by k times equation i.

Type III Elementary Operation: **add to one equation a multiple of another equation**

$\left.\begin{array}{l} e_i \to e_i + c\, e_j \\ i \neq j,\, c \in \mathbb{F} \end{array}\right\}$ Replace equation i by adding to it a multiple of equation j.

Remarks

- You could argue that the interchange of two equations is a linear type of process.
- Scaling e_i by a constant, k, is linear: $k\, e_i$ is a linear combination of the elements in the set $\{e_i\}$.
- The construction of $e_i \to e_i + c\, e_j$, $(i \neq j,\, c \in \mathbb{F})$, is a linear combination of the elements in the set $\{e_i, e_j\}$.

Thus our three allowed operations are linear operations on the original linear system, and they yield a new (equivalent) linear system.

Warning

In practice there might be other operations which you want to perform. These operations will not be Elementary Operations, instead they will be **combinations** of elementary operations, for example:

$$e_i \to ce_i + de_j,\ \ c \in \mathbb{F}\setminus\{0,1\}\,, d \in \mathbb{F}\setminus\{0\}\,.$$

As is traditional in a linear algebra course, we will **not** make use of these non Elementary Operations. However, there will be occasions when it would be efficient to do so, and we did use this step to solve Example 2 (7A) (this refers to Example 2 in the notes Topics 7A: we will use this notation throughout the course.)

We will manipulate any system $(*)$ making use of the three Elementary Operations only.

- If at any point, we produce an equation of the form

$$0 = a,\ \text{where } a \neq 0,$$

then the system is inconsistent and we stop at once.

- In the following analysis we will assume that we have a consistent system.

Definition 9: Trivial equation

We refer to the equation $0 = 0$, as the **trivial equation**.

Any other equation is known as a **non-trivial** equation.

The trivial equation is always true, and it has no information content.

- If at any point we produce the trivial equation when performing Elementary Operations, then we can move this equation to the end of the system, by performing a type I Elementary Operation. You might be tempted to ignore such equations, and not write them down. In practice, we keep any (there may be more than one) equations of this form in our new system, so that we will always have the same number of equations in any of the equivalent systems.

The goal of performing the elementary operations is to obtain a simpler system, that is, one whose solution set is easy to obtain. Let us refer to this as the, "final equivalent system." There is no reason why your final equivalent system and my final equivalent system would be identical.

Example 11

Consider the system of three linear equations in three unknowns:

$$\begin{cases} -2x_1 - 4x_2 - 6x_3 = 4 & (e_1) \\ 3x_1 + 6x_2 + 10x_3 = 6 & (e_2) \\ x_2 + 2x_3 = 5 & (e_3) \end{cases}$$

Suppose that I do the Elementary Operation:

$$e_1 \to -\tfrac{1}{2}\, e_1 \quad \text{to get the new system} \quad \begin{cases} x_1 + 2x_2 + 3x_3 = -2 & (e_1) \\ 3x_1 + 6x_2 + 10x_3 = 6 & (e_2) \\ x_2 + 2x_3 = 5 & (e_3) \end{cases}$$

Notice that we use the same labels for the new system. That is in this case, the new equation one is labelled again e_1.
I have done this so that the first equation tells me about 1 times x_1. I will now make use of this new equation 1, to remove any x_1 terms from the original equation (e_2).

$$e_2 \to e_2 - 3e_1 \quad \text{gives} \quad \begin{cases} x_1 + 2x_2 + 3x_3 = -2 & (e_1) \\ 0x_1 + 0x_2 + 1x_3 = 12 & (e_2) \\ x_2 + 2x_3 = 5 & (e_3) \end{cases}$$

This could be **your final equivalent system.** The second equation (e_2) tells you that $x_3 = 12$. You could make use of this in the third equation (e_3) to get $x_2 + 2(12) = 5$, so that $x_2 = -19$. You can put the values for x_2 and x_3 into equation (e_1), to get: $x_1 + 2(-19) + 3(12) = -2$, so that $x_1 = 0$.

On the other hand, I may wish to continue applying elementary operations. I perform the operation

$$e_2 \leftrightarrow e_3 \quad \text{to get} \quad \begin{cases} x_1 + 2x_2 + 3x_3 = -2 & (e_1) \\ x_2 + 2x_3 = 5 & (e_2) \\ x_3 = 12 & (e_3) \end{cases}$$

so that the second equation (e_2) has information about the second variable. Then I perform:

$$e_2 \to e_2 - 2e_3 \quad \text{to obtain} \quad \begin{cases} x_1 + 2x_2 + 3x_3 = -2 & (e_1) \\ x_2 + 0x_3 = -19 & (e_2) \\ x_3 = 12 & (e_3) \end{cases}$$

$$e_1 \to e_1 - 3e_3 \quad \text{to get} \begin{cases} x_1 + 2x_2 + 0x_3 = -38 & (e_1) \\ x_2 + 0x_3 = -19 & (e_2) \\ x_3 = 12 & (e_3) \end{cases}$$

$$e_1 \to e_1 - 2e_2 \;\text{ to finally obtain }\; \begin{cases} x_1 + 0x_2 + 0x_3 = 0 & (e_1) \\ \quad\; x_2 + 0x_3 = -19 & (e_2) \\ \qquad\quad x_3 = 12 & (e_3) \end{cases}$$

This is now **my final equivalent system**, since I can just read the solution set from it as:

$$S = \{\mathbf{x} \in \mathbb{R}^3 : x_1 = 0,\; x_2 = -19,\; x_3 = 12\}.$$

It is clear from this example that there will be plenty of choices on which elementary operations you apply at any stage, and when you could stop at your final equivalent system.

There is not a, "best," final equivalent system, however there are two particular appearances that we like the final equivalent system to have.

We will attempt to motivate these two types of appearance, for the final equivalent system, and then discuss how they can be obtained.

Let us consider a system of m linear equations and n unknowns.
We normally try to solve for the variables of the system in the alphabetical or numerical order in which they appear. Let us assume that the variables are labelled $x_1, x_2, \cdots, x_n$.
We assume that there is at least one equation with the variable x_1 appearing in it with a non-zero coefficient. It would be rather strange if the variable x_1 did not appear in any of the equations, and if this were the case then you could always relabel the variables.

We would like the first equation to tell us about the first variable, x_1, so that it is of the form,

$$k_1\, x_1 + \cdots = c_1,$$

with $k_1 \neq 0$. We may choose to scale k_1 to 1, although this is not necessary.
We now consider the variable x_1 as solved for in the (new) first equation.
We will make use of this equation to remove the variable x_1 from all the other equations **after** the first equation, using type III elementary operations.

We would like the second equation to tell us about the next variable (alphabetically or numerically), that can be obtained, call this x_q. Often this would be the variable x_2, so that $q = 2$, but this is not always the case. The (new) second equation has the form:

$$k_q\, x_q + \cdots = c_2,$$

with $k_q \neq 0$. We may choose to scale k_q to 1, although this is not necessary.
We now consider the variable x_q as solved for in the (new) second equation.
We will make use of this equation to remove the variable x_q from all the other equations, **after** the second equation, using type III elementary operations.

Repeat this process, moving down through the equations, and solving for the next variable (alphabetically or numerically) that it is possible to solve for, and being aware that this may require you to interchange the order of some of the equations, i.e. perform type I elementary operations. Also move any trivial equation(s) to the end of the system.

$\vdots$

etc.

$\vdots$

We assume that the t^{th} equation is the last non trivial equation, and it is possible that t is equal to m. We also assume that this equation tells us about the r^{th} variable, x_r: note this equation should not involve the variables which the previous equations have been solved for, e.g. x_1 from equation 1, x_q from equation 2, etc.

$$k_r\, x_r + \cdots = c_t,$$

with $k_r \neq 0$. We may choose to scale k_r to 1, although this is not necessary.
We now consider the variable x_r as solved for in the (new) t^{th} equation.
It might happen that $x_r = x_m$, that is, you solve the t^{th} equation for the m^{th} variable, but this will not always be the case.

This form of the equivalent system of equations has a **particularly simple appearance**. We may choose to stop performing elementary operations at this point, and so this would be our **final equivalent system**. We will refer to this new system as system ($**$).

Note that my system ($**$) and your system ($**$) need not be the same. For example, the scaling factors in our equations (the k_1, k_2, $\cdots$, k_r above) are not unique. That is, this final equivalent system is **not** unique.

Definition 10: Forward elimination

Proceeding from ($*$) to ($**$) is known as the **forward elimination** phase.

Definition 11: Back substitution

Once the forward elimination phase is complete, we can obtain the solution set using a process known as **back substitution.** This is the name given to the procedure in which we use the information given in the t^{th} equation for the variable x_r to then **substitute** for the variable x_r into all the previous equations in your final equivalent system. Then we repeat this process moving **back** up the system. With the last step being to use the information given in the 2^{nd} equation for the variable x_q to **substitute** for the variable x_q into the first equation (if it appears there).

An alternative way to proceed, instead of performing back substitution, is to continue to simplify the system. The following extension yields a unique final equivalent system, which is particularly simple.

First of all, scale each of the non trivial equations in the current form of the system, equation 1 the factor $\frac{1}{k_1}$, equation 2 by the factor $\frac{1}{k_q}, \cdots$, equation t by the factor $\frac{1}{k_r}$, i.e., perform type II elementary operations.

We now make use of the t^{th} equation to eliminate the variable x_r from all the equations **before (above)** equation e_r.
This is accomplished by using type III elementary operations.

We then examine e_{r-1}, which we assume is of the form,

$$(1)\, x_s + \cdots = c_{r-1},$$

for some real s. This equation tells us what the variable x_s is. We now make use of this equation to remove the variable x_s from all the equations before (above) equation e_{r-1}. This is accomplished by using more type III elementary operations.

Repeat this process, moving up through the equations, and performing elementary type III operations to remove the variables which have been solved for from the equations above them.

$\vdots$

etc.

$\vdots$

We conclude by using equation 2 to remove the variable x_q from first equation, e_1.

Definition 12: Backward elimination

These additional steps, performed after the end of the forward elimination phase, are collectively known as the **backward elimination phase**. When we are performing these steps we say that we are doing backward elimination.

We will refer to this final equivalent system as system $(***)$.

The system $(***)$ is unique, that means that you and I will have the exact same final equivalent system if we perform both the forward and backward elimination phases.

This will be true even when our systems $(**)$ are not identical.

Topic 7C

Systems of Equations

(Continuation of Topic 7B)

Example 12

Consider the system of 5 equations in 4 unknowns:

$$(*) \quad \begin{cases} \begin{aligned} x_1 + 2x_2 &= 1 \quad (e_1) \\ x_1 + 2x_2 + 3x_3 + x_4 &= 0 \quad (e_2) \\ -x_1 - x_2 + x_3 + x_4 &= -2 \quad (e_3) \\ x_2 + x_3 + x_4 &= -1 \quad (e_4) \\ -x_2 + 2x_3 &= 0 \quad (e_5) \end{aligned} \end{cases}$$

We will solve this system. We will apply elementary operations, usually just one or two at a time. At each stage, we will produce an equivalent system. We will stop the process when we produce an equivalent system whose solution set is relatively easily obtained.

We notice that the first equation has x_1 in it and we will use it to eliminate the variable x_1 from all the equations below the first equation.

If the first equation did not have any x_1 terms then we would have switched it with one that did. It would be very strange if x_1 did not appear in any of the equations.

We perform,

$$e_2 \to e_2 - e_1$$
$$e_3 \to e_3 + e_1$$

We notice that the fourth and fifth equations do not contain x_1, and so they are not modified at this stage. We get,

$$\begin{cases} \begin{aligned} x_1 + 2x_2 &= 1 \quad (e_1) \\ 3x_3 + x_4 &= -1 \quad (e_2) \\ x_2 + x_3 + x_4 &= -1 \quad (e_3) \\ x_2 + x_3 + x_4 &= -1 \quad (e_4) \\ -x_2 + 2x_3 &= 0 \quad (e_5) \end{aligned} \end{cases}$$

Notice that once we have determined our elementary operations and we have performed them then we **use the same equation labels** for the new system. e.g. the equation labelled (e_2) in the system above is a different equation from that labelled (e_2) in the original system.

We now notice that the second equation does not have an x_2 variable, but the third one does. It seems like a good idea to perform the elementary operation:

$$e_2 \leftrightarrow e_3,$$

to get,

$$\begin{cases} x_1 + 2x_2 = \quad 1 & (e_1) \\ x_2 + x_3 + x_4 = -1 & (e_2) \\ 3x_3 + x_4 = -1 & (e_3) \\ x_2 + x_3 + x_4 = -1 & (e_4) \\ -x_2 + 2x_3 = \quad 0 & (e_5) \end{cases}$$

The new second equation has x_2 in it and we can make use of this to remove the variable x_2 from the equations below e_2.

$$e_4 \to e_4 - e_2$$

$$e_5 \to e_5 + e_2$$

yields:

$$\begin{cases} x_1 + 2x_2 = \quad 1 & (e_1) \\ x_2 + x_3 + x_4 = -1 & (e_2) \\ 3x_3 + x_4 = -1 & (e_3) \\ 0 = \quad 0 & (e_4) \\ 3x_3 + x_4 = -1 & (e_5) \end{cases}$$

The third equation has x_3 in it, which we like, and we make use of it to remove x_3 from all the equations below it. In this case we need only perform:

$$e_5 \to e_5 - e_3,$$

to obtain,

$$(**) \quad \begin{cases} x_1 + 2x_2 = \quad 1 & (e_1) \\ x_2 + x_3 + x_4 = -1 & (e_2) \\ 3x_3 + x_4 = -1 & (e_3) \\ 0 = \quad 0 & (e_4) \\ 0 = \quad 0 & (e_5) \end{cases}$$

Notice that the two trivial equations are at the end.

We have now completed the forward elimination stage.
This system is a good choice for a final equivalent system.

We can obtain the solution of both this final equivalent system, and of the original system, by **back substitution**. To perform back substitution, we start at the last equation, and solve it. We then look at the second last equation and solve this one. We proceed **back up** the system, solving each equation in the system, one by one, as we **substitute** for the variables that we have already solved for.

e_5 is satisfied.

e_4 is satisfied.

Notice that the third equation as telling us about the variable x_3 and x_4. This means that one of these variables can "arbitrarily" be assigned any real value, and that the value of the other variable will depend on this arbitrary variable. Let us give them names.

Definition 13: Free variable

We say that an unknown variable is a **free variable** when we have the choice to assign it any possible real value.

Definition 14: Basic variable

A variable which is not a free variable, is called a **basic variable**.

If a system has one or more free variables, i.e when we are have to assign any scalar value to one or more variables, then we (usually can) make a choice which unknown(s) is/are free variable(s). It is traditional, but not necessary, to solve for the basic variables in decreasing numerical or alphabetical order: we will do so. We will then be able to use the system of equations to solve for any basic variable in terms of constants and free variables.

Returning to Example 12, we will let x_4 be the free variable, thus, the variable x_3 will be a basic variable. If we let $x_4 = t \in \mathbb{R}$, then

e_3 tells us that $x_3 = \frac{1}{3}(-1 - t)$.

e_2 tells us that $x_2 = -1 - x_3 - x_4 = -1 - \frac{1}{3}(-1 - t) - t = -\frac{2}{3} - \frac{2t}{3}$.

e_1 tells us that $x_1 = 1 - 2x_2 = 1 - 2(-\frac{2}{3} - \frac{2t}{3}) = \frac{7}{3} + \frac{4t}{3}$.

The solution set is thus

$$S = \left\{ \begin{pmatrix} \frac{7}{3} + \frac{4\,t}{3} \\ -\frac{2}{3} - \frac{2\,t}{3} \\ -\frac{1}{3} - \frac{t}{3} \\ t \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Alternatively, if we do not wish to do the back substitution, we can continue performing elementary operations.

We return to system $(**)$, in which we observe that the last two equations are trivially satisfied, and that the third equation is in terms of both variables x_3 and x_4. Following the tradition, we then let the variable x_4 (the one with the highest numerical order), be a free variable. We then scale the third equation by $\frac{1}{3}$, yielding:

$$e_3 \to \frac{1}{3}e_3,$$

$$\begin{cases} x_1 + 2x_2 = \ \ 1 & (e_1) \\ x_2 + x_3 + x_4 = -1 & (e_2) \\ x_3 + \frac{1}{3}x_4 = \frac{-1}{3} & (e_3) \\ 0 = \ \ 0 & (e_4) \\ 0 = \ \ 0 & (e_5) \end{cases}$$

The third equation has been solved for the basic variable x_3. We make use of the third equation to eliminate the basic variable x_3 from the equations above it, by performing

$$e_2 \to e_2 - e_3, \quad \text{to get}$$

$$\begin{cases} x_1 + 2x_2 = \ \ 1 & (e_1) \\ x_2 + \frac{2}{3}x_4 = \frac{-2}{3} & (e_2) \\ x_3 + \frac{1}{3}x_4 = \frac{-1}{3} & (e_3) \\ 0 = \ \ 0 & (e_4) \\ 0 = \ \ 0 & (e_5) \end{cases}$$

The second equation determines the second basic variable x_2, and so we use it to eliminate x_2 from e_1.

$$e_1 \to e_1 - 2e_2, \quad \text{to get}$$

$$(***) \quad \begin{cases} x_1 - \frac{4}{3}x_4 = \frac{7}{3} & (e_1) \\ x_2 + \frac{2}{3}x_4 = \frac{-2}{3} & (e_2) \\ x_3 + \frac{1}{3}x_4 = \frac{-1}{3} & (e_3) \\ 0 = \ \ 0 & (e_4) \\ 0 = \ \ 0 & (e_5) \end{cases}$$

The first equation determines the basic variable x_1.

There is no equation which determines x_4, since it is the free variable. We notice that:

e_1 will tell us what $(1)\, x_1$ is in terms of x_4.

e_2 will tell us what $(1)\, x_2$ is in terms of x_4.

e_3 will tell us what $(1)\, x_3$ is in terms of x_4.

The final equivalent system $(***)$ is the unique system which determines the variables $(1)\, x_1$, $(1)\, x_2$ and $(1)\, x_3$ in that order if we let $x_4 = s \in \mathbb{R}$: we then solve for the basic variables, i.e.

e_3 tells us that $x_3 = \frac{1}{3}(-1 - s)$,

e_2 tells us that $x_2 = -\frac{2}{3}(1 + s)$, and

e_1 tells us that $x_1 = \frac{7}{3} + \frac{4}{3}s$.

The solution set is thus

$$S = \left\{ \begin{pmatrix} \frac{7}{3} + \frac{4s}{3} \\ -\frac{2}{3} - \frac{2s}{3} \\ -\frac{1}{3} - \frac{s}{3} \\ s \end{pmatrix} : s \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{pmatrix} + s \begin{pmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{pmatrix} : s \in \mathbb{R} \right\}.$$

If we used another parameter, $u = \dfrac{s}{3}$, then we remove some of the fractions, to get,

$$S = \left\{ \begin{pmatrix} \frac{7}{3} + 4u \\ -\frac{2}{3} - 2u \\ -\frac{1}{3} - u \\ 3u \end{pmatrix} : u \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{pmatrix} + u \begin{pmatrix} 4 \\ -2 \\ -1 \\ 3 \end{pmatrix} : u \in \mathbb{R} \right\}.$$

Alternatively, we could let $p = \dfrac{s-2}{3}$, and get an even nicer looking solution set S:

$$S = \left\{ \begin{pmatrix} 5 \\ -2 \\ -1 \\ 2 \end{pmatrix} + p \begin{pmatrix} 4 \\ -2 \\ -1 \\ 3 \end{pmatrix} : p \in \mathbb{R} \right\}.$$

Topic 8A

Systems of Equations using Matrices

When we are solving the system of equations, we manipulate the system using elementary operations. At each stage we have an equivalent system of equations, that is, each system has the same solution set, and you can stop at any stage, with any equivalent system, at which you think that you can write down the solution set.

We notice that we have written down the variables x_1, x_2, x_3, and x_4 many many times during the processing of the problem, however they are only really important at the very beginning and at the very end.

Moving forwards, we will only need to write down the **coefficients** in the intermediate steps.

Thus, if we have the system $(*)$,

$$(*) \quad \left\{ \begin{array}{cc} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 & (e_1) \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 & (e_2) \\ \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m & (e_m) \end{array} \right.$$

then we introduce the coefficient matrix.

Definition 1: Coefficient matrix

We define the **coefficient matrix**, A, of the system to mean the array,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1(n-1)} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2(n-1)} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{(m-1)1} & a_{(m-1)2} & \cdots & a_{(m-1)(n-1)} & a_{(m-1)n} \\ a_{m1} & a_{m2} & \cdots & a_{m(n-1)} & a_{mn} \end{pmatrix},$$

which is built up from the coefficients in the system of equations.

The coefficient matrix, A, has m rows and n columns.

Definition 2: Augmented matrix

We define the **augmented matrix**, $B = (A|\mathbf{b})$, of the system (*) to mean the array,

$$B = (A|\mathbf{b}) = \left(\begin{array}{ccccc|c} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n-1} & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{(m-1)1} & a_{(m-1)2} & \cdots & a_{(m-1)(n-1)} & a_{(m-1)n} & b_{m-1} \\ a_{m1} & a_{m2} & \cdots & a_{m(n-1)} & a_{mn} & b_m \end{array} \right).$$

The augmented matrix, $B = (A|\mathbf{b})$, has m rows and $n+1$ columns.

Clearly the augmented matrix is just the coefficient matrix A with an extra column added, namely the column corresponding to the right hand side of the system of equations, and so we write column vector $\mathbf{b}$.

In solving the system of equations, we can just manipulate the rows of the augmented matrix, instead of carrying with us the additional redundant variables.

Definition 3:

Let C be a matrix with p rows and q columns. We refer to the number in the i^{th} row and j^{th} column of C as the $(i,j)^{th}$ entry of C and label it c_{ij} or $(C)_{ij}$.

If A is a coefficient matrix of a system of equations, then a_{ij} is the coefficient of the variable x_j in the i^{th} equation.

If B is the augmented matrix of this system, then $b_{ij} = a_{ij}$ is the coefficient of x_j in the i^{th} equation if $1 \leq j \leq n$, and $b_{i(n+1)} = b_i = (\mathbf{b})_i$, which is the constant term on the RHS of the i^{th} equation of the system. Read this again!

Definition 4: Elementary Row Operation (ERO)

The operations which correspond to the elementary operations on the equations are called **elementary row operations** when they are performed on the coefficient and/or augmented matrix.

Elementary operation	Equation	Row
Type I	$e_i \leftrightarrow e_j$	$R_i \leftrightarrow R_j$
Type II	$e_i \to me_i \ , \quad m \neq 0$	$R_i \to mR_i \ , \quad m \neq 0$
Type III	$e_i \to e_i + me_j \ , \quad i \neq j$	$R_i \to R_i + mR_j \ , \quad i \neq j$

At any point in the process of manipulating the rows of the augmented matrix, you can stop, and immediately write down the corresponding system of equations.

The first column in the array corresponds to the variable x_1,

the second column in the array corresponds to the variable $x_2, \cdots$

$\cdots$ the second-last column in the array corresponds to the variable x_n,

the dashed line corresponds to the equals sign,

and the last column in the array corresponds to the constants on the right hand side (RHS) of the original system.

Example 1:

Let us re-consider Example 1 (7A), and solve it using elementary row operations. We have the system:

$$(*) \quad \begin{cases} x_1 + 2x_2 = 1 & (e_1) \\ x_1 + 2x_2 + 3x_3 + x_4 = 0 & (e_2) \\ -x_1 - x_2 + x_3 + x_4 = -2 & (e_3) \\ x_2 + x_3 + x_4 = -1 & (e_4) \\ -x_2 + 2x_3 = 0 & (e_5) \end{cases}$$

We will now solve this system using the same elementary row operations as before, however this time we will make use of matrices.

The augmented matrix for this system is,

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ 1 & 2 & 3 & 1 & 0 \\ -1 & -1 & 1 & 1 & -2 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & -1 & 2 & 0 & 0 \end{array} \right).$$

The first row contains information about x_1 and we still proceed by thinking that we are eliminating x_1 from the other equations. We perform the following elementary row operations:

$$R_2 \to R_2 - R_1$$
$$R_3 \to R_3 + R_1$$

to get

$$\left(\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & -1 & 2 & 0 & 0 \end{array}\right).$$

The second row does not contain information about the variable x_2, but the third one does. So we switch these two rows, $R_2 \leftrightarrow R_3$ and get:

$$\left(\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & -1 & 2 & 0 & 0 \end{array}\right).$$

We now make use of the new second row to eliminate the x_2 variable from all equations **after** the second one, by performing the following EROs:

$$R_4 \to R_4 - R_2$$
$$R_5 \to R_5 + R_2$$

$$\left(\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & -1 \end{array}\right).$$

The third row tells us about x_3, and we use this to eliminate x_3 from the equations **after** the third, by performing the following ERO:

$$R_5 \to R_5 - R_3,$$

$$(**) \quad \left(\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This is the first point at which we could stop the forward elimination phase.
This point is one of the intelligent ones at which to stop. Notice that this augmented matrix, and the accompanying coefficient matrix have a particularly simple structure.

Definition 5: Zero row

We refer to a row in a matrix that has all of its entries as zeros, as a **zero row**.

In an augmented matrix, a zero row corresponds to the equation: $0 = 0$, and is not a cause for concern.

In this example, the last two rows of the augmented matrix are zero rows. They have no information content. We do still write them down in any further calculation(s) so that we do not forget that we did start with 5, and not 3, equations.

Note however, if we had a zero row in a coefficient matrix, and the last term in the same corresponding row in the augmented matrix is $c \neq 0$, then we can deduce that our system is inconsistent since one of the equations becomes: $0 = c$, where $c \neq 0$.

A system of equations that is equivalent to the original system $(*)$ is:

$$\begin{cases} x_1 + 2x_2 = 1 & (e_1) \\ x_2 + x_3 + x_4 = -1 & (e_2) \\ 3x_3 + x_4 = -1 & (e_3) \\ 0 = 0 & (e_4) \\ 0 = 0 & (e_5) \end{cases}$$

which may be solved by back substitution as follows.

Let $x_4 = t \in \mathbb{R}$, then (e_3) tells us that $x_3 = \frac{1}{3}(-1 - t)$.

(e_2) tells us that $x_2 = -1 - x_3 - x_4 = -1 - \frac{1}{3}(-1 - t) - t = -\frac{2}{3} - \frac{2\,t}{3}$

$(e_{1)}$ tells us that $x_1 = 1 - 2x_2 = 1 - 2(-\frac{2}{3} - \frac{2\,t}{3}) = \frac{7}{3} + \frac{4\,t}{3}$.

The solution set is

$$S = \left\{ \begin{pmatrix} \frac{7}{3} + \frac{4\,t}{3} \\ -\frac{2}{3} - \frac{2\,t}{3} \\ -\frac{1}{3} - \frac{t}{3} \\ t \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{pmatrix} + t \begin{pmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Or we can continue with our elementary row operations.

The last two rows are both zero rows, and these both correspond to the consistent equation $0 = 0$.

The third row has information about the variable x_3. We scale it so that it tells us about $(1)\, x_3$ (and not $3\, x_3$), by performing the following ERO:

$$R_3 \to \frac{1}{3} R_3,$$

to get

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

We then use row 3 to eliminate x_3 from the equation(s) above it, in this case just the second one, by performing this time, the following ERO:

$$R_2 \to R_2 - R_3$$

yielding,

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The last step of the backward phase is to eliminate x_2 from the first equation, which is achieved by the ERO:

$$R_1 \to R_1 - 2R_2$$

yielding,

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{4}{3} & \frac{7}{3} \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

We stop the backward elimination phase at this point.
There is no more simplification possible, or required.

Earlier we remarked on the simple structure of the (reduced) augmented and coefficients matrices at the step labelled (∗∗). Notice that every matrix since that stage has had a similar simple structure. You could argue that the last one obtained is the simplest of them all. We will introduce terminology to refer to these structures in the next few lectures.

The system of equations corresponding to the last augmented matrix above is:

$$\begin{cases} x_1 - \frac{4}{3}x_4 = \frac{7}{3} & (e_1) \\ x_2 + \frac{2}{3}x_4 = -\frac{2}{3} & (e_2) \\ x_3 + \frac{1}{3}x_4 = -\frac{1}{3} & (e_3) \\ 0 = 0 & (e_4) \\ 0 = 0 & (e_5) \end{cases}$$

At this point we can write down the solution.

Note x_4 is a free variable and thus we can let $x_4 = t \in \mathbb{R}$, and then,

e_3 tells us that $x_3 = \frac{1}{3}(-1 - t)$,

e_2 tells us that $x_2 = -\frac{2}{3}(1 + t)$, and

e_1 tells us that $x_1 = \frac{7}{3} + \frac{4}{3}t$.

And the solution set is

$$S = \left\{ \begin{pmatrix} \frac{7}{3} + \frac{4t}{3} \\ -\frac{2}{3} - \frac{2t}{3} \\ -\frac{1}{3} - \frac{t}{3} \\ t \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{pmatrix} + t \begin{pmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Topic 8B
The Gauss-Jordan Algorithm

Suppose that you have a system of linear equations, and you wish to obtain the solution set. There are many ways in which you could attempt to do this, for example you could begin by re-arranging them in your favorite order; you could then choose to solve the first equation for a variable which you pick; then examine the second equation, and manipulate it in order to solve it for the second variable of your choice, etc.

The Gauss-Jordan algorithm provides us with one particular strategy to use for solving the system of equations, and it is the formal way of referring to the strategy that we have already introduced in Topic 8A. It assumes that we have already agreed on some preferred ordering of the variables. There is still some freedom in some of the steps that you take, and/or the order in which you carry them out.

In addition to explaining the algorithm, we will be introducing some language and notation in this section: it is crucial that you learn to use this language as soon as possible.

We have labelled the original system of equations as system $(*)$. We will also label the augmented matrix corresponding to system $(*)$ with a $(*)$ as well.

It is important to remember that if at any point during our manipulations of the augmented matrix, when we have a new augmented matrix labelled M, say, we can pause and write down the system of equations which corresponds to this augmented matrix M. *We will refer to this system as the corresponding system.* The corresponding system to M will be equivalent to the original system $(*)$. If $1 \leq i \leq m$ and $1 \leq j \leq n$, then the $(i, j)^{th}$ entry of M, denoted by m_{ij}, will provide the coefficient of the variable x_j in the i^{th} equation in this equivalent system of equations.

We assume that:

a) we have a system of m linear equations in n unknowns.

b) the equations are labelled in a specific order, $e_1, e_2, \cdots, e_m$.

c) the (unknown) variables are labelled as $x_1, x_2, \cdots, x_n$, and we wish to solve the system for the variables in this order.

d) the variable x_1 appears in at least one of the equations, i.e. the $(i, 1)^{th}$ entry of the augmented matrix $(*)$ is non-zero for some $i,\ 1 \le i \le m$. If there were no x_1's in any of the equations, then the system provides no information about x_1, and it is thus a free variable (use a parameter for it): we could relabel the system with the old x_2 being the new x_1, etc.

We will address the issue of the matrix form of the system of equations, that is, we will be referring to the augmented matrix of the system.

Note: **not all texts agree on the exact form of the algorithm.**

The Gauss Part of the Algorithm.

Step 1, aim: to obtain a non-zero entry in the augmented matrix in the $(1,1)$ position.

In system $(*)$ choose **any** row, the p^{th} say, which has a non-zero term in the first column. If $p \ne 1$, then perform a type I ERO: $R_1 \leftrightarrow R_p$. The $(1,1)^{th}$ entry in the (new) augmented matrix is non-zero.

The (new) first equation, in the corresponding system, tells us about the variable x_1.

We will make use of the (new) $(1,1)$ entry to manipulate the augmented matrix so that all entries in column 1 (except the $(1,1)$ entry) are zero. This is equivalent to making the coefficients of x_1 in all the equations after the first equation of the corresponding system, equal to zero.
Clearly, this $(1,1)$ entry is important, and it is given a name.

The next definitions apply to both the coefficient matrix and the augmented matrix of a system of equations, and we will indicate this with coefficient/augmented.

Definition 6a

We say that column 1 of the coefficient/augmented matrix is a **pivot column.**

We say that the $(1,1)$ position in the coefficient/augmented matrix is a **pivot position.**

We call any **non-zero** entry in the $(1,1)$ position of the coefficient/augmented matrix, a **pivot** .

To be precise, these are examples of a pivot column, pivot position, and a pivot.
Definition 6a is extended to subsequent rows and columns other than the first row and first column.
Definition 6b, at the end of this topic, provides the formal definitions of these terms.

Step 2, aim: to transform all entries in column 1, after the first row, to zero.

Perform type III EROs to make all the $(i, 1)$ entries zero, with $1 < i \leq m$.

There will be occasions when we perform this operation one step at a time, but we usually perform these type III EROs in groups of more than one.

We will change neither the first row nor the first column of the augmented matrix in any of the *following* manipulations using the *Gauss* part of the algorithm.

Step 3, aim: to obtain a non-zero entry in the augmented matrix in the $(2, q)^{th}$ position, where $(q > 1)$ is the smallest integer for which this is possible.

Examine the columns of the (new) augmented matrix, except column 1.
Find the smallest value of q $(q > 1)$ such that the $(i, q)^{th}$ entry of the matrix (where $1 < i \leq m$) is non-zero.

If there is no such q, then you only have one non trivial equation, in the system corresponding to the (new) augmented matrix. This is the first equation, and there is no further simplification possible using the Gauss algorithm.

Choose **one** row, the k^{th} say, which has a non-zero term in the q^{th} column. If $k \neq 2$, then perform a type I ERO: $R_2 \leftrightarrow R_k$. The $(2, q)^{th}$ entry in the (new) augmented matrix is non-zero.

The q^{th} column is a pivot column, the $(2, q)$ position is a pivot position, and the $(2, q)^{th}$ entry is a pivot.

The second equation, in the corresponding system, tells us about the variable x_q.

Step 4, aim: to transform all entries in column q, after the second row, to zero.

Perform type III EROs to make all the $(i, q)^{th}$ entries zero, with $2 < i \leq m$.

Step 5, repeat: move down the system of equations, repeating the procedure.

If the (a, b) position is a pivot position, then the subsequent pivot position is in the closest possible column to the right of column b, and in the closest possible row below row a. A single type I ERO may be required to achieve this.

The Gauss part of the algorithm is now completed.

Notice that at this point, any zero rows will appear at the bottom of your new augmented matrix, and will not lie above any non-zero rows.

Definition 7: Row echelon form.

We say that a matrix is in **row echelon form** to mean that:

(i) all zero rows occur as the last rows in the matrix.

(ii) the first entry, from the LHS, in any non-zero row appears to the right of the first entry in any rows above it.

Definition 8: Row echelon form of a matrix, A, denoted by $REF(A)$.

Let A be a matrix, we say that the matrix C is a **row echelon form** of A , written $REF(A)$, to mean that C is in row echelon form and that C has been obtained from A by performing a finite number of elementary row operations to A.

The Gauss part of the algorithm ends when the augmented matrix has been manipulated into a row echelon form.

Note: If A is not a zero matrix (a matrix for which all the entries are zero), then there are an infinite number of matrices which are row echelon forms of A.

Once the Gauss part of the algorithm (also called **Gaussian elimination**) is complete, you may write down the new equivalent system of equations and solve it by back substitution.

Note: if the last column of the ***augmented matrix*** *is a pivot column, then you can stop immediately as the system is inconsistent , and has no solutions, because one of the equations is* $0 = 1$.

The Jordan Part of the Algorithm

The Jordan part of the algorithm has two pieces.
The **first piece** of the Jordan part of the algorithm is to scale all the pivots to 1.

Step 6, aim: make all the pivots 1.

Use type II EROs to scale all the pivots to 1.

Definition 9: Leading entry and leading variable

The first non-zero entry from the LHS in any row of a matrix is called a **leading entry**.

If a leading entry lies in column k, then we refer to the variable, x_k, as a **leading variable.**

Examples: the variables x_1 and x_q are leading variables in our general algorithm.

Definition 10: Leading 1

We use the term, **leading 1**, to refer to a leading entry which is the number 1.

Example: a pivot that has been scaled to unity is the most important example of a leading one. If the $(p, q)^{th}$ position is a pivot position and the pivot is a leading 1, then we think of the p^{th} equation, in the corresponding system of equations, as telling us about $(1)\, x_q$.

Note: some texts do not consider the Gauss part of the algorithm to be complete until you have scaled all the pivots to unity, i.e. until all the leading entries in a row echelon form of the matrix are unity.

The **second piece** of the Jordan part of the algorithm aims to manipulate the augmented matrix so that the only non-zero entry in a pivot column is the pivot itself.

We continue by examining the pivot which is the furthest to the right. We assume that the pivot position is the $(t, r)^{th}$ position, so that the pivot is in the t^{th} row, and the pivot column is the r^{th} column.

The r^{th} entry in the t^{th} row is a leading entry and x_r is a leading variable.

The t^{th} equation, in the corresponding system, tells us about the variable x_r.

Note that if $t = 1$, then you can stop as there is nothing further to do since there is only one non-zero equation.

Note that if $r = n + 1$, then there is a pivot in the last column of *the augmented matrix,* and then *the system is inconsistent.* In this case you can stop. Otherwise continue.

Step 7, aim: to transform all entries in column r, except the one in the t^{th} row, to zero.

Perform type III EROs to make all the $(i,r)^{th}$ entries, a zero, with $1 \leq i \leq m, i \neq t$. Note that if $i > t$, then the $(i,r)^{th}$ entries are already zero.

Once this has been done, then the variable x_r has been removed from all the equations, in the corresponding system, except from the t^{th} equation.

This concludes Step 7. Before going to the next step, let us assume that the $(t-1,s)^{th}$ position is a pivot position, that is, x_s is a leading variable.

Step 8, aim: to transform all entries in column s, except the one in the $(t-1)^{th}$ row, to zero.

Perform type III EROs to make all the $(i,s)^{th}$ entries zero, with $1 \leq i \leq m, i \neq (t-1)$. Note that if $i > (t-1)$, then the $(i,s)^{th}$ entries are already zero.

Step 9, aim: to transform the entries in any pivot column to zero, except for the row containing that pivot.

Repeat this process for the pivots in rows $(t-2), (t-3), \cdots, 2$, in this order.

You have now completed the Jordan part of the algorithm.

You have now completed the **backward-elimination** part of the algorithm.

At this point, the system of equations cannot be simplified any further.

Definition 11: Reduced row echelon form.

We say that a matrix is in **reduced row echelon form** (RREF) to mean that:

(i) it is in echelon form.

(ii) all the pivots are one.

(iii) the only non-zero entry in a pivot column is the pivot itself.

Thus, once we have completed the Jordan part of the algorithm, then our augmented matrix will be in reduced row echelon form.

Definition 12: Reduced row echelon form of a matrix, A, denoted $RREF(A)$

Let A be a matrix, we say that the matrix R is the reduced row echelon form of A , written $RREF(A)$, to mean that R is the matrix in reduced row echelon form that has been obtained from A by performing a finite number of elementary row operations to A.

Lemma 1

If A is a matrix, then there is a **unique** matrix R such that $R = RREF(A)$.

Definition 6b: Pivot column, pivot position and pivot.

Let $A \in M_{m\times n}(\mathbb{F})$, let the matrix C be a REF(A), and let R be the unique $RREF(A)$.

Let (i, j) be the position of a leading entry (which must be a leading 1 in R).

(i) We say that the column j of the matrix C is a **pivot column** of A.

(ii) We say that the $(i, j)^{\text{th}}$ position of the matrix C is a **pivot position** of A.

(iii) If the $(i, j)^{\text{th}}$entry in the matrix C is non-zero, then we refer to that $(i, j)^{\text{th}}$ entry as a **pivot** of A.

(iv) We say that the column j of the matrix C is a **pivot column** of C.

(v) We say that the $(i, j)^{\text{th}}$ position of the matrix C is a **pivot position** of C.

(vi) If the $(i, j)^{\text{th}}$entry in the matrix C is non-zero, then we refer to that $(i, j)^{\text{th}}$ entry as a **pivot** of C.

Topic 8C

Condensed Gauss-Jordan Algorithm

We assume that:

a) we have a system of m linear equations in n unknowns.

b) the equations are labelled in a specific order, $e_1, e_2, \cdots, e_m$.

c) the (unknowns) variables are labelled as $x_1, x_2, \cdots, x_n$, and we wish to solve the system for the variables in this order.

d) the variable x_1, appears in at least one of the equations, equivalently, we assume that the first column of the augmented matrix is not a column of zeros.

We will refer to the augmented matrix corresponding to the system here, and label the augmented matrix corresponding to system $(*)$ as augmented matrix $(*)$ too.

Elementary Row Operation is abbreviated by ERO.

Step 1
In the augmented matrix $(*)$, choose **any** row which has a non-zero entry in column 1.
If this is not already row 1, then perform a type I EROs to make it the new row 1.

Step 2
Perform type III EROs to make all the $(i, 1)$ entries zero, for $1 < i \leq m$.

Step 3
Perform a single ERO of type I, if necessary, so that the $(2, q)^{th}$ entry of the augmented matrix is non-zero, and q $(q > 1)$ is the smallest integer for which this is possible.
If this is not possible then you have already completed the Gauss (part of the) algorithm.

Step 4
Perform type III EROs to make all the $(i, q)^{th}$ entries zero, with $2 < i \leq m$.

Step 5: Move all the way down the system of equations, repeating this procedure.

The Gauss part of the algorithm is now completed.

The Jordan part of the algorithm will now begin.

Step 6: In each non-zero row, use type II EROs to scale the leading entry to be 1, a leading 1.

Suppose that the $(t, r)^{th}$ pivot position is furthest to the right.
If $r = n + 1$, then stop as the system is inconsistent.

Step 7: Perform type III EROs to make all of the $(i, r)^{th}$ entries zero, with $1 \leq i \neq t \leq m$.

Step 8: Repeat this process for all the pivots in rows $(t - 1), (t - 2), \cdots, 2$, in this order.

Example 2:

Let us now re-consider Example 1 (T8A), and solve it using the Gauss-Jordan algorithm.

$$(*) \quad \left\{ \begin{array}{rrl} x_1 + 2x_2 = & 1 & (e_1) \\ x_1 + 2x_2 + 3x_3 + x_4 = & 0 & (e_2) \\ -x_1 - x_2 + x_3 + x_4 = & -2 & (e_3) \\ x_2 + x_3 + x_4 = & -1 & (e_4) \\ -x_2 + 2x_3 = & 0 & (e_5) \end{array} \right.$$

The augmented matrix for this system is

$$\begin{pmatrix} 1 & 2 & 0 & 0 & | \, 1 \\ 1 & 2 & 3 & 1 & | \, 0 \\ -1 & -1 & 1 & 1 & | -2 \\ 0 & 1 & 1 & 1 & | -1 \\ 0 & -1 & 2 & 0 & | \, 0 \end{pmatrix}$$

Let us process the matrix using the Gauss part of the algorithm.

$$R_2 \to R_2 - R_1$$
$$R_3 \to R_3 + R_1$$

$$\begin{pmatrix} 1 & 2 & 0 & 0 & | \, 1 \\ 0 & 0 & 3 & 1 & | -1 \\ 0 & 1 & 1 & 1 & | -1 \\ 0 & 1 & 1 & 1 & | -1 \\ 0 & -1 & 2 & 0 & | \, 0 \end{pmatrix}$$

$$R_2 \leftrightarrow R_3$$

$$\begin{pmatrix} 1 & 2 & 0 & 0 & | \, 1 \\ 0 & 1 & 1 & 1 & | \, -1 \\ 0 & 0 & 3 & 1 & | \, -1 \\ 0 & 1 & 1 & 1 & | \, -1 \\ 0 & -1 & 2 & 0 & | \, 0 \end{pmatrix}$$

$$R_4 \to R_4 - R_2$$
$$R_5 \to R_5 + R_2$$

$$\begin{pmatrix} 1 & 2 & 0 & 0 & | \, 1 \\ 0 & 1 & 1 & 1 & | \, -1 \\ 0 & 0 & 3 & 1 & | \, -1 \\ 0 & 0 & 0 & 0 & | \, 0 \\ 0 & 0 & 3 & 1 & | \, -1 \end{pmatrix}$$

$$R_5 \to R_5 - R_3,$$

$$(**) \qquad \left\{ \begin{pmatrix} 1 & 2 & 0 & 0 & | \, 1 \\ 0 & 1 & 1 & 1 & | \, -1 \\ 0 & 0 & 3 & 1 & | \, -1 \\ 0 & 0 & 0 & 0 & | \, 0 \\ 0 & 0 & 0 & 0 & | \, 0 \end{pmatrix} \right.$$

The Gauss algorithm is complete.
The forward elimination process is complete.
The solution of the system may be obtained by back substitution.
The above matrix is in **row echelon form**.

Let us now continue to process the matrix using the Jordan part of the algorithm.

$$R_3 \to \frac{1}{3} R_3,$$

$$\begin{pmatrix} 1 & 2 & 0 & 0 & | \, 1 \\ 0 & 1 & 1 & 1 & | \, -1 \\ 0 & 0 & 1 & \frac{1}{3} & | \, -\frac{1}{3} \\ 0 & 0 & 0 & 0 & | \, 0 \\ 0 & 0 & 0 & 0 & | \, 0 \end{pmatrix}.$$

The above matrix is also in row echelon form.

$$R_2 \to R_2 - R_3,$$

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The above matrix is also in row echelon form.

$$R_1 \to R_1 - 2R_2,$$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{4}{3} & \frac{7}{3} \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The above matrix is now in reduced row echelon form.

This is the end of the Gauss-Jordan Algorithm.
The solution can be written down with little effort by solving for each variable in any order without any need for back-substitution.
Be careful of the signs of terms and of the parameter(s).

We suggests that you actually write down the corresponding system of equations, at least for your first few examples.

The solution set is:

$$S = \left\{ \begin{pmatrix} \frac{7}{3} + \frac{4t}{3} \\ -\frac{2}{3} - \frac{2t}{3} \\ -\frac{1}{3} - \frac{t}{3} \\ t \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{pmatrix} + t \begin{pmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Topic 8D
Canonical Gauss-Jordan

If you choose to solve a system of equations using either the Gauss or the Gauss-Jordan algorithm, then there is usually considerable choice in which operations you perform, and when you perform them. If you compare your solution set with a friend, and it is identical, then we are not concerned with how we obtained the solution. However, if you do not have identical solution sets, then it is often difficult to determine which one of you has made a mistake, especially if you did not perform identical steps in your calculation.

In the version of the Gauss-Jordan algorithm which I present in this lecture, there are **no choices** to be made at any stage.
We will refer to this method as the **Canonical Gauss-Jordan** method, and we will **usually** use this method to solve systems of equations in these notes, **and** any solutions that we write out for assignments. If you use this method, then you will be able to compare your solution(s) with ours, and thus easily identify the position of an error, if there is one.

There are two additional constraints in this method.

> The first constraint is that once we start to work with a pivot, that is, in its pivot position, then we **immediately scale the pivot to be unity**: i.e. we make the pivot a leading 1 as soon as we start to make use of it.
>
> The second constraint is the following. Suppose that we have a pivot in the position (i, j), and we have manipulated the augmented matrix so that there are zeros in column j, after the i^{th} row. Once we have identified the next pivot column, then we must select the non-zero entry in the **first row below** the i^{th} row as the pivot in that pivot column.

Here is the condensed Canonical Gauss-Jordan Algorithm.

We assume that:

a) we have a system of linear equations.

b) the equations are labelled in a specific order, $(e_1), (e_2), \cdots, (e_m)$.

c) the (unknowns) variables are labelled as $x_1, x_2, \cdots, x_n$, and we wish to solve the system. for the variables in this order.

d) the variable x_1, appears in at least one of the equations.

Step 1, aim: to obtain a pivot in the $(1, 1)$ position.
In system $(*)$ choose **the first** row which has a non-zero term in the column 1.
If necessary perform a type I ERO, to obtain a pivot in the $(1, 1)$ position.

Step 2 (new), aim: to obtain a leading 1 in the $(1, 1)$ position.
If necessary, scale row 1 with a type II ERO so that the pivot in this row is unity.

Step 3, aim: to transform all entries in column 1, after row 1, to zero.
Perform type III EROs to make all the $(i, 1)$ entries zero, for $1 < i \leq m$.

We will change neither row 1 nor column 1 of the augmented matrix in any of the *following* manipulations using the *Gauss* part of the algorithm.

Step 4, aim: to obtain a pivot in the augmented matrix in the $(2, q)^{th}$ position, where q $(q > 1)$ is the smallest integer for which this is possible. If a type I ERO is required to do this, then we select the closest row below row 2. If this is not possible, then there is no further simplification possible using the Gauss algorithm.

Examine the columns of the (new) augmented matrix, except column 1.
Find the smallest value of q $(q > 1)$ such that the $(i, q)^{th}$ entry of the matrix (where $1 < i \leq m$) is non-zero.
Choose the smallest row index after the first, the k^{th} say, which has a non-zero term in the q^{th} column.
If $k \neq 2$ then perform a type I ERO: $R_2 \leftrightarrow R_k$. The $(2, q)^{th}$ entry in the (new) augmented matrix is (now) a pivot.

Step 5 (new), aim: to obtain a leading 1 in the $(2, q)$ position.
If necessary, scale row two with a type II operation so that the pivot in this row is unity.

Step 6, aim: to transform all entries in column q, after row 2, to zero.
Perform type III EROs to make all the (i, q) entries zero, for $2 < i \leq m$.

Step 7, Repeat: move down the system of equations, repeating the procedure.
If the (a, b) position is a pivot position then the subsequent pivot position is in the closest possible column to the right of column b, in the closest possible row below row a.
A single type I ERO may be required to achieve this. If necessary, scale the new pivot to unity.

We assume that the pivot, which is the furthest to the right, is in a pivot position $(t, r)^{th}$.

Step 8 (new), aim: to obtain a leading 1 in the $(t, r)^{th}$ pivot position.
Scale the t^{th} row with a type II ERO so that the $(t, r)^{th}$ pivot is unity.

The Gauss part of the algorithm is now completed.

The Jordan part of the algorithm now begins.

Note - the old step 6 is no longer needed as the pivots are already unity.

Step 9, aim: to transform all entries in column r, except that in the t^{th} row, to zero. Perform type III EROs to make all the $(i,r)^{th}$ entries zero, with $1 \le i \le m, i \ne t$.

We assume that $((t-1,s))$ is a pivot position, that is, x_s is a leading entry.

Step 10, aim: to transform all entries in column s, except that in the $(t-1)^{th}$ row, to zero. Perform type III EROs to make all the $(i,s)^{th}$ entries zero, with $1 \le i \le m, i \ne (t-1)$.

Step 11: Repeat this process for the pivots in rows $(t-2),(t-3),\cdots,2$, in this order.

Example 1:

Consider the system of 4 equations in 4 unknowns:

$$\begin{cases} 3x_1 - 4x_2 - 1x_3 - 19x_4 = -8 & (e_1) \\ 2x_1 - 3x_2 + x_3 - 22x_4 = -1 & (e_2) \\ 1x_1 + 2x_2 - x_3 + 7x_4 = 2 & (e_3) \\ 6x_1 - 12x_2 + 2x_3 - 70x_4 = -12 & (e_4) \end{cases}$$

The augmented matrix for this system is

$$\left(\begin{array}{rrrr|l} 3 & -4 & -1 & -19 & -8 \\ 2 & -3 & 1 & -22 & -1 \\ 1 & 2 & -1 & 7 & 2 \\ 6 & -12 & 2 & -70 & -12 \end{array}\right).$$

We will solve this system of equations several different ways.

Solution 1

First of all we will just perform the Gauss (and Gauss-Jordan) algorithms, stating the elementary row operations, but without any additional explanation. This is the way in which you will solve systems once you have mastered the techniques. In solution 2, which you may wish to look over first, we apply the same steps in the same order, but also provide more explanation. In solution 3, we use the canonical Gauss-Jordan algorithm.

$$R_1 \longleftrightarrow R_3 \quad \text{yields}$$

$$\left(\begin{array}{rrrr|r} 1 & 2 & -1 & 7 & 2 \\ 2 & -3 & 1 & -22 & -1 \\ 3 & -4 & -1 & -19 & -8 \\ 6 & -12 & 2 & -70 & -12 \end{array}\right) \quad \text{and then,}$$

$$\begin{aligned} R_2 &\to R_2 - 2R_1 \quad \text{and} \\ R_3 &\to R_3 - 3R_1 \quad \text{and} \\ R_4 &\to R_4 - 6R_1 \quad \text{give} \end{aligned}$$

$$\left(\begin{array}{rrrr|r} 1 & 2 & -1 & 7 & 2 \\ 0 & -7 & 3 & -36 & -5 \\ 0 & -10 & 2 & -40 & -14 \\ 0 & -24 & 8 & -112 & -24 \end{array}\right). \quad \text{Then}$$

$$\begin{aligned} R_3 &\to R_3 - \frac{10}{7}R_2 \quad \text{and} \\ R_4 &\to R_4 - \frac{24}{7}R_2 \quad \text{give} \end{aligned}$$

$$\left(\begin{array}{rrrr|r} 1 & 2 & -1 & 7 & 2 \\ 0 & -7 & 3 & -36 & -5 \\ 0 & 0 & \frac{-16}{7} & \frac{80}{7} & -\frac{48}{7} \\ 0 & 0 & \frac{-16}{7} & \frac{80}{7} & -\frac{48}{7} \end{array}\right), \quad \text{and}$$

$$R_4 \to R_4 - R_3 \quad \text{produces}$$

$$\left(\begin{array}{rrrr|r} 1 & 2 & -1 & 7 & 2 \\ 0 & -7 & 3 & -36 & -5 \\ 0 & 0 & \frac{-16}{7} & \frac{80}{7} & -\frac{48}{7} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

$$\begin{aligned} R_2 &\to -\frac{1}{7}R_2 \quad \text{and} \\ R_3 &\to -\frac{7}{16}R_3 \quad \text{give} \end{aligned}$$

$$\left(\begin{array}{rrrr|r} 1 & 2 & -1 & 7 & 2 \\ 0 & 1 & \frac{-3}{7} & \frac{36}{7} & \frac{5}{7} \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right), \quad \text{and then}$$

$$\begin{aligned} &R_2 \to R_2 + \frac{3}{7}R_3 \quad \text{and} \\ &R_1 \to R_1 + R_3 \qquad \text{yield} \end{aligned}$$

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 2 & 5 \\ 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right). \text{ Finally}$$

$$R_1 \to R_1 - 2R_2 \quad \text{produces:}$$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -4 & 1 \\ 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Both the coefficient matrix and the augmented matrix are in Reduced Row Echelon Form. An equivalent system of equations, to the original system, is:

$$\begin{cases} x_1 - 4x_4 = 1 & (e_1) \\ x_2 + 3x_4 = 2 & (e_2) \\ x_3 - 5x_4 = 3 & (e_3) \\ \qquad\quad 0 = 0 & (e_4) \end{cases}$$

If we let $x_4 = t,\ t \in \mathbb{R}$, then (e_3) tells us that $x_3 = 3 + 5t$.

(e_2) tells us that $x_2 = 2 - 3t$, and

(e_1) tells us that $x_1 = 1 + 4t$.

The solution set is

$$S = \left\{ \begin{pmatrix} 1+4t \\ 2-3t \\ 3+5t \\ t \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} + t \begin{pmatrix} 4 \\ -3 \\ 5 \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Solution 2

We will follow the same steps as above, but we will also explain the steps and make use of some of the new language.

We see that there are some x_1 terms in this system. Thus x_1 is a basic variable, column 1 is a pivot column, and the $(1, 1)$ position is a pivot position. At the moment there is a 3 in the pivot position, we have various choices. We can leave the pivot as a 3 or change its value by interchanging the first row with another row which has a non-zero entry in column 1.

We choose to interchange rows 1 and 3 so that the pivot is 1, a leading 1: it might be easier to work with.

$$R_1 \longleftrightarrow R_3 \quad \text{yields}$$

$$\left(\begin{array}{rrrr|l} 1 & 2 & -1 & 7 & 2 \\ 2 & -3 & 1 & -22 & -1 \\ 3 & -4 & -1 & -19 & -8 \\ 6 & -12 & 2 & -70 & -12 \end{array}\right).$$

We now make use of this pivot in the $(1, 1)$ position to obtain entries of zero in column 1, after the first row. Equivalently, in the corresponding system of equations, we remove the x_1 term from all equations after the first. This is accomplished by performing type III EROs.

$$\begin{aligned} R_2 &\to R_2 - 2R_1 \text{ and} \\ R_3 &\to R_3 - 3R_1 \text{ and} \\ R_4 &\to R_4 - 6R_1 \text{ give} \end{aligned}$$

$$\left(\begin{array}{rrrr|l} 1 & 2 & -1 & 7 & 2 \\ 0 & -7 & 3 & -36 & -5 \\ 0 & -10 & 2 & -40 & -14 \\ 0 & -24 & 8 & -112 & -24 \end{array}\right).$$

We can ignore the first column for the rest of the calculation.

We observe that the second column is also a pivot column since at least one row after the first contains a non-zero entry. The pivot in the $(2, 2)$ pivot position is -7, and we make use of this pivot to set all entries in column 2 to zero, after the second row.

Equivalently, in the corresponding system of equations, we remove the x_2 term from all equations after the second.

$$\begin{aligned} R_3 &\to R_3 - \frac{10}{7}R_2 \text{ and} \\ R_4 &\to R_4 - \frac{24}{7}R_2 \text{ give} \end{aligned}$$

$$\left(\begin{array}{rrrr|l} 1 & 2 & -1 & 7 & 2 \\ 0 & -7 & 3 & -36 & -5 \\ 0 & 0 & \frac{-16}{7} & \frac{80}{7} & -\frac{48}{7} \\ 0 & 0 & \frac{-16}{7} & \frac{80}{7} & -\frac{48}{7} \end{array}\right).$$

We observe that the third column is a pivot column since at least one row after the second contains a non-zero entry. The pivot in the $(3,3)$ pivot position is $\frac{-16}{7}$, and we make use of this pivot to set all entries in column 3 to zero, after the third row. This is achieved by applying a single type III ERO.

Equivalently, in the corresponding system of equations, we remove the x_3 term from all equations after the third.

$$R_4 \to R_4 - R_3 \text{ produces}$$

$$\left(\begin{array}{rrrr|l} 1 & 2 & -1 & 7 & 2 \\ 0 & -7 & 3 & -36 & -5 \\ 0 & 0 & \frac{-16}{7} & \frac{80}{7} & -\frac{48}{7} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This matrix (both the coefficient and the augmented) is in row echelon form. We observe that the corresponding system of equations is consistent. The fourth column is not a pivot column and so the variable x_4, in the corresponding system, is a free variable.

The Gauss algorithm is complete, and we could stop here and obtain the solution using back substitution.

We will continue with the Jordan part of the algorithm.

We first scale all of the pivots to unity.

$$R_2 \to -\frac{1}{7}R_2 \quad \text{and}$$
$$R_3 \to -\frac{7}{16}R_3 \text{ give}$$

$$\left(\begin{array}{rrrr|l} 1 & 2 & -1 & 7 & 2 \\ 0 & 1 & \frac{-3}{7} & \frac{36}{7} & \frac{5}{7} \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

For the corresponding system of equations:

The first equation tells us about $1x_1$.

The second equation tells us about $1x_2$.

The third equation tells us about $1x_3$.

The fourth equation is the identity, $0 = 0$, which is called the **trivial equation**.

We do not have an equation for x_4, this variable is a free variable.

We make use of the pivot in the third row to obtain entries of zeros in column 3, above the third row by performing type III EROs. Equivalently, in the corresponding system of equations, we remove the x_3 term from all equations above the third.

$$\begin{aligned} &R_2 \to R_2 + \frac{3}{7}R_3 \quad \text{and} \\ &R_1 \to R_1 - R_3 \qquad \text{yield} \end{aligned}$$

$$\begin{pmatrix} 1 & 2 & 0 & 2 & | \, 5 \\ 0 & 1 & 0 & 3 & | \, 2 \\ 0 & 0 & 1 & -5 & | \, 3 \\ 0 & 0 & 0 & 0 & | \, 0 \end{pmatrix}.$$

We make use of the pivot in the second row to obtain entries of zeros in column 2, above the second row by performing type III EROs. Equivalently, in the corresponding system of equations, we remove the x_2 term from all equations above the second.

$$R_1 \to R_1 - 2R_2 \ \text{ produces :}$$

$$\begin{pmatrix} 1 & 0 & 0 & -4 & | \, 1 \\ 0 & 1 & 0 & 3 & | \, 2 \\ 0 & 0 & 1 & -5 & | \, 3 \\ 0 & 0 & 0 & 0 & | \, 0 \end{pmatrix}.$$

Notice both coefficient matrix and the augmented matrix are in Reduced Row Echelon Form. An equivalent system of equation, to the original system, is:

$$\begin{cases} x_1 - 4x_4 = 1 & (e_1) \\ x_2 + 3x_4 = 2 & (e_2) \\ x_3 - 5x_4 = 3 & (e_3) \\ \qquad\quad 0 = 0 & (e_4) \end{cases}$$

If we let $x_4 = t \in \mathbb{R}$, then (e_3) tells us that $x_3 = 3 + 5t$.

(e_2) tells us that $x_2 = 2 - 3t$

(e_1) tells us that $x_1 = 1 + 4t$.

The solution set is

$$S = \left\{ \begin{pmatrix} 1+4t \\ 2-3t \\ 3+5t \\ t \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} + t \begin{pmatrix} 4 \\ -3 \\ 5 \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Solution 3

We will now solve this system using the Canonical Gauss-Jordan Algorithm.

There is already a pivot in the $(1,1)$ position: let us scale it to unity.

$$R_1 \to \frac{1}{3} R_1$$

$$\left(\begin{array}{rrrr|l} 1 & \frac{-4}{3} & \frac{-1}{3} & \frac{-19}{3} & \frac{-8}{3} \\ 2 & -3 & 1 & -22 & -1 \\ 1 & 2 & -1 & 7 & 2 \\ 6 & -12 & 2 & -70 & -12 \end{array} \right)$$

Use the leading one in the $(1,1)$ position to obtain entries of zero in column 1, after row 1.

$$\begin{array}{ll} R_2 \to R_2 - 2R_1 & \text{and} \\ R_3 \to R_3 - R_1 & \text{and} \\ R_4 \to R_4 - 6R_1 & \text{give} \end{array}$$

$$\left(\begin{array}{rrrr|l} 1 & \frac{-4}{3} & \frac{-1}{3} & \frac{-19}{3} & \frac{-8}{3} \\ 0 & \frac{-1}{3} & \frac{5}{3} & \frac{-28}{3} & \frac{13}{3} \\ 0 & \frac{10}{3} & \frac{-2}{3} & \frac{40}{3} & \frac{14}{3} \\ 0 & -4 & 4 & -32 & 4 \end{array} \right).$$

There is already a pivot in the $(2,2)$ position: let us scale it to unity.

$$R_2 \to -3R_2$$

$$\left(\begin{array}{cccc|c} 1 & \frac{-4}{3} & \frac{-1}{3} & \frac{-19}{3} & \frac{-8}{3} \\ 0 & 1 & -5 & 28 & -13 \\ 0 & \frac{10}{3} & \frac{-2}{3} & \frac{40}{3} & \frac{14}{3} \\ 0 & -4 & 4 & -32 & 4 \end{array}\right).$$

Use the leading one in the $(2,2)$ position to obtain entries of zero in column 2, after row 2.

$$R_3 \to R_3 - \frac{10}{3}R_2$$
$$R_4 \to R_4 + 4R_2$$

$$\left(\begin{array}{cccc|c} 1 & \frac{-4}{3} & \frac{-1}{3} & \frac{-19}{3} & \frac{-8}{3} \\ 0 & 1 & -5 & 28 & -13 \\ 0 & 0 & 16 & -80 & 48 \\ 0 & 0 & -16 & 80 & -48 \end{array}\right).$$

There is already a pivot in the $(3,3)$ position: let us scale it to unity.

$$R_3 \to \frac{1}{16}R_3$$

$$\left(\begin{array}{cccc|c} 1 & \frac{-4}{3} & \frac{-1}{3} & \frac{-19}{3} & \frac{-8}{3} \\ 0 & 1 & -5 & 28 & -13 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & -16 & 80 & -48 \end{array}\right).$$

Use the leading one in the $(3,3)$ position to obtain entries of zero in column 3, after row 3.

$$R_4 \to R_4 + 16R_3$$

$$\left(\begin{array}{cccc|c} 1 & \frac{-4}{3} & \frac{-1}{3} & \frac{-19}{3} & \frac{-8}{3} \\ 0 & 1 & -5 & 28 & -13 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Both the coefficient and the augmented matrices are now in row echelon form.
We observe that the corresponding system of equations is consistent.
The fourth column is not a pivot column and so the variable x_4 is a free variable.

The Gauss algorithm is complete, and we could stop here and obtain the solution using back substitution.

We proceed with the Jordan algorithm.

We use the third row to obtain entries of zero in column 3 above row 3.

$$R_2 \to R_2 + 5R_3 \quad \text{and}$$
$$R_1 \to R_1 + \frac{1}{3}R_3 \quad \text{give}$$

$$\left(\begin{array}{cccc|c} 1 & \frac{-4}{3} & 0 & -8 & \frac{-5}{3} \\ 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

We use the second row to obtain entries of zero in column 2 above row 2.

$$R_1 \to R_1 + \frac{4}{3}R_2$$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -4 & 1 \\ 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 1 & -5 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Both the coefficient and the augmented matrices are in Reduced Row Echelon Form. An equivalent system of equation, to the original system, is:

$$\begin{cases} x_1 - 4x_4 = 1 & (e_1) \\ x_2 + 3x_4 = 2 & (e_2) \\ x_3 - 5x_4 = 3 & (e_3) \\ \qquad\quad 0 = 0 & (e_4) \end{cases}$$

If we let $x_4 = t \in \mathbb{R}$, then (e_3) tells us that $x_3 = 3 + 5t$.

(e_2) tells us that $x_2 = 2 - 3t$

(e_1) tells us that $x_1 = 1 + 4t$.

The solution set is

$$\left\{ \begin{pmatrix} 1+4t \\ 2-3t \\ 3+5t \\ t \end{pmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} + t \begin{pmatrix} 4 \\ -3 \\ 5 \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Topic 9

Some Language and Some Counting

Let us begin with a system of m linear equations in n unknowns.

We can peel off the essential information content from this system to yield the coefficient matrix, A, which will have m rows and n columns, and the augmented matrix, $B = (A|\mathbf{b})$, which has m rows and $(n+1)$ columns.

Notation 1: $M_{p\times q}(\mathbb{R})$, $M_{p\times q}(\mathbb{C})$, $M_{p\times q}$.

We use the notation $M_{p\times q}(\mathbb{R})$ to denote the set of all matrices with p rows, q columns, and whose entries are all real numbers. We use the notation $M_{p\times q}(\mathbb{C})$ to denote the set of all matrices with p rows, q columns, and whose entries are all complex numbers. When we do not need to distinguish whether we are dealing with real or complex numbers then we abbreviate these to $M_{p\times q}$. We read this as the set of p by q matrices.

In the system $(*)$, $A \in M_{m\times n}$ and $B \in M_{m\times(n+1)}$, read as A is an m by n matrix and B is an m by $(n+1)$ matrix.

Some of the important questions relating to our system concern the solution set.

I) Is the system consistent or not?
II) If the system is consistent, then is there a unique solution?
III) If the system is consistent, and if there is not a unique solution, then how many parameters are there in the solution ?

We will be able to answer these questions once we have completed the Gauss algorithm, and before we have actually found the solution set.

Definition 1*:* Rank

Let $C \in M_{p\times q}$, we say that C has **rank** r, or the **rank** of C is r, written $rank(C) = r$, to mean that there will be exactly r pivots when C is manipulated into row echelon form.

The rank of a matrix provides important information about that matrix.

Notice that since there is at most one pivot in each row, then $rank(C) = r \leq p$, also, since there is at most one pivot in each column, then $rank(C) = r \leq q$.

Notice that if $A \in M_{m\times n}$, and if $rank(A) = m$, then all the rows of the matrix A have a pivot in them. Since the matrix $(A|\mathbf{b})$ does not have any additional rows to A, it cannot have any additional pivots. In this case $rank(A) = rank(A|\mathbf{b})$ and we conclude that the system of equations is consistent.

Lemma 1

The system of linear equations *is consistent* **iff** $rank(A) = rank(A|\mathbf{b})$.

Proof

We recall that the two matrices A and $(A|\mathbf{b})$ have the same number number of rows, and that the only difference between them is that $(A|\mathbf{b})$ has an additional last column of $\mathbf{b}$. Thus there are only two possibilities:

$$\textbf{either} \;\; rank(A|\mathbf{b}) = rank(A) \;\; \textbf{or} \;\; rank(A|\mathbf{b}) = rank(A) + 1.$$

Suppose that we perform identical elementary row operations on the two matrices A and $(A|\mathbf{b})$ until we have manipulated A into row echelon form, and then stop. We will be able to determine $rank(A)$ and $rank(A|\mathbf{b})$ at this stage: **either**

(a) $rank(A) = rank(A|\mathbf{b})$, and the system is consistent. You could now solve it by back substitution, **or**

(b) $rank(A) \neq rank(A|\mathbf{b})$, and the last column of $(A|\mathbf{b})$ is a pivot column. In this case, the last non-zero row of the manipulated augmented matrix corresponds to the equation $0 = 1$, and the system is inconsistent. ■

Example 1

Suppose we have a linear system of equations and we perform elementary row operations on the augmented matrix and obtain the row echelon form:

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

Clearly the rank of the coefficient matrix is 3 and the rank of the augmented matrix is 4, and we conclude that the system of equations is inconsistent since the last equation is $0 = 1$.

Example 2

Suppose we have a linear system of equations and we perform elementary row operations on the augmented matrix and obtain the row echelon form:

$$\left(\begin{array}{cccc|c} 2 & -5 & 5 & -7 & 4 \\ 0 & 0 & 4 & -9 & 3 \\ 0 & 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Clearly the ranks of both the coefficient matrix and of the augmented matrix is 3. We conclude that the system of equations is consistent.

We now have a simple procedure which will allow us to determine whether or not our linear system of equations is consistent: just compare the two ranks: $rank(A)$, and $rank(A|\mathbf{b})$.

If you discover that your system is inconsistent, then you are usually finished with it.

If your system is consistent then there are solutions, and you probably want to know how many parameters there are, if any, in the solution set. This is also determined from $rank(A)$.

Definition 2: Nullity

Let $C \in M_{p\times q}$, with $rank(C) = r$, we define the nullity of C, written, $nullity(C)$ to mean the number $q - r$.

Lemma 2: Number of parameters in the solution set (Rank-Nullity Theorem Part-I)

If the system of linear equations $(*)$ *is consistent,* where $A \in M_{m\times n}$ and $rank(A) = r$, then the solution set to this system will contain $n - r$ parameters.

Thus the nullity of A tells us the number of parameters that there will be in the solution set to a consistent system of equations with A as the coefficient matrix.

Proof

Suppose that we perform elementary row operations on the augmented matrix $(A|\mathbf{b})$ and produced a row echelon form of $(A|\mathbf{b})$, with corresponding equivalent system $(**)$.

Suppose that the $(i,j)^{th}$ position of the augmented matrix $(A|\mathbf{b})$ is a pivot position. This means that the i^{th} equation in $(**)$ tells us what the variable x_j is equal to.

Since $rank(A) = rank(A|\mathbf{b}) = r$, there are r pivots, and thus, there are r non-zero equations in the system $(**)$. The system $(**)$ determines each one of the variables corresponding to the pivot columns. These r variables are the basic variables.

There is no more information in the system $(**)$, as all the equations after the r^{th}, if any, are $0 = 0$. All the other variables, if any, are free variables.

Since we commenced with a system in n variables, and there are r basic variables which are completely determined by the system then the remaining $n - r$ variables are not determined by the system, there are $n - r$ free variables in the system and thus exactly $n - r$ different parameters in the solutions set. $\blacksquare$

In order to obtain the solution set, you may either perform back substitution after Gauss elimination or proceed to complete the Jordan part of the algorithm, and then write down the solution set.

Example 3 (Example 2 continued)

The row echelon form of the augmented matrix in Example 2 is:

$$\left(\begin{array}{cccc|c} 2 & -5 & 5 & -7 & 4 \\ 0 & 0 & 4 & -9 & 3 \\ 0 & 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The rank of the coefficient matrix is equal to the rank of the augmented matrix: the system is thus consistent. There are 4 variables and 3 pivots.
Using Lemma 2, we conclude there will be $4 - 3 = 1$ parameter in the solution set to the system.
The variable x_2 is the (only) free variable, since only column 2 of the coefficient matrix is not a pivot column.

Topic 10A

More Real Examples

Example 1

Consider the system of linear equations:

$$\begin{cases} x - 2y - z + 3w = 1 \\ 2x - 4y + z = 5 \\ x - 2y + 2z - 3w = 4 \end{cases}$$

The augmented matrix is :

$$\left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 2 & -4 & 1 & 0 & 5 \\ 1 & -2 & 2 & -3 & 4 \end{array}\right).$$

Let us perform the following elementary row operations using Canonical Gauss-Jordan:

$$\begin{cases} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - R_1 \end{cases} \quad \text{to get} \quad \left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -6 & 3 \end{array}\right).$$

$$R_2 \to \tfrac{1}{3}R_2 \quad \text{yields} \quad \left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 3 & -6 & 3 \end{array}\right).$$

$$R_3 \to R_3 - 3R_2 \quad \text{yields} \quad \left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right),$$

This matrix is in row echelon form, and since $rank(A) = rank(A|\mathbf{b}) = 2$, then we conclude that **this system is consistent.**

Let us now perform the ERO,

$$R_1 \to R_1 + R_2 \quad \text{to get the RREF} \quad \left(\begin{array}{rrrr|r} 1 & -2 & 0 & 1 & 2 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

There are 4 unknowns and 2 pivots: we thus need to introduce two free variables. We can let $y = s$ and $w = t$, where $s, t \in \mathbb{R}$.

Solving for the basic variables, we get: $x = 2 + 2s - t$ and $z = 1 + 2t$, for $s, t \in \mathbb{R}$. We can then write the solution set, S as

$$S = \left\{ \begin{pmatrix} 2+2s-t \\ s \\ 1+2t \\ t \end{pmatrix} : \ s,t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : \ s,t \in \mathbb{R} \right\}.$$

Example 2

Let us now consider the system of linear equations:

$$\begin{cases} x - 2y - z + 3w = 1 \\ 2x - 4y + z = 2 \\ x - 2y + 2z - 3w = 3 \end{cases}$$

The augmented matrix is:

$$\left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 2 & -4 & 1 & 0 & 2 \\ 1 & -2 & 2 & -3 & 3 \end{array} \right).$$

You should notice that the coefficient matrix for this system is exactly the same as it was for the previous system, so we will be performing the exact same steps in our row reduction, the only changes will be in the last column of the augmented matrix.

$$\begin{cases} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - R_1 \end{cases} \text{ yield } \left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 3 & -6 & 0 \\ 0 & 0 & 3 & -6 & 2 \end{array} \right).$$

$$R_2 \to \tfrac{1}{3}R_2 \text{ to get } \left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 3 & -6 & 2 \end{array} \right).$$

$$R_3 \to R_3 - 3R_2 \text{ to obtain } \left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{array} \right).$$

This matrix is in row echelon form, and this time, since $rank(A) = 2$ is not equal to $rank(A|\mathbf{b}) = 3$, we now conclude that **this system is inconsistent**.

The solution set S is then $S = \emptyset$.

Example 3

Let us now consider the third system of linear equations:

$$\begin{cases} x - 2y - z + 3w = -1 \\ 2x - 4y + z = 4 \\ x - 2y + 2z - 3w = 5 \end{cases}$$

The augmented matrix is:

$$\left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & -1 \\ 2 & -4 & 1 & 0 & 4 \\ 1 & -2 & 2 & -3 & 5 \end{array}\right).$$

Once again, you should observe that we have the same coefficient matrix, and thus we will be performing the same steps as in the two previous examples:

$$\begin{cases} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - R_1 \end{cases} \text{ to get } \left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & -1 \\ 0 & 0 & 3 & -6 & 6 \\ 0 & 0 & 3 & -6 & 6 \end{array}\right).$$

$$R_2 \to \tfrac{1}{3}R_2 \text{ yields } \left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & -1 \\ 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 3 & -6 & 6 \end{array}\right).$$

$$R_3 \to R_3 - 3R_2 \text{ to obtain } \left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & -1 \\ 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This matrix is in row echelon form, and again, since $rank(A) = rank(A|\mathbf{b}) = 2$, we then conclude that **this system is consistent.**

Let us now perform this ERO,

$$R_1 \to R_1 + R_2 \text{ to get the RREF } \left(\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 1 \\ 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

There are again 4 unknowns and 2 pivots, and thus two free variables.
We can let $y = a$ and $w = b$, where $a, b \in \mathbb{R}$.

Solving for the basic variables, we get (this time) $x = 1 + 2a - b$ and $z = 2 + 2b$, for $a, b \in \mathbb{R}$.
We can write the solution set, T as

$$T = \left\{ \begin{pmatrix} 1 + 2a - b \\ a \\ 2 + 2b \\ b \end{pmatrix} : a, b \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + a \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

Example 4

Lastly, let us now consider the system of linear equations:

$$\begin{cases} x - 2y - z + 3w = 0 \\ 2x - 4y + z = 0 \\ x - 2y + 2z - 3w = 0 \end{cases}$$

The augmented matrix is :

$$\left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 0 \\ 2 & -4 & 1 & 0 & 0 \\ 1 & -2 & 2 & -3 & 0 \end{array}\right),$$

Once again, this system has the same coefficient matrix as the other three examples. You could argue that this is the simplest system to solve since all the terms on the RHS are zero. We will proceed as before.

$$\begin{cases} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - R_1 \end{cases} \text{ to get } \left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & -6 & 0 \\ 0 & 0 & 3 & -6 & 0 \end{array}\right).$$

$$R_2 \to \tfrac{1}{3}R_2 \text{ yields } \left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 3 & -6 & 0 \end{array}\right).$$

$$R_3 \to R_3 - 3R_2 \text{ to obtain } \left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This matrix is in row echelon form, and since $rank(A) = rank(A|\mathbf{b}) = 2$, we then also conclude that **this system is consistent.**

Let us now perform this ERO,

$$R_1 \to R_1 + R_2 \text{ to get the RREF } \left(\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Again, there are 4 unknowns and 2 pivots, and thus two free variables.
We can let $y = p$ and $w = q$, where $p, q \in \mathbb{R}$.

Solving for the basic variables, we get (this time): $x = 2p - q$ and $z = 2q$, for $p, q \in \mathbb{R}$.
We can write the solution set, U as

$$U = \left\{ \begin{pmatrix} 2p - q \\ p \\ 2q \\ q \end{pmatrix} : p, q \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + p \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + q \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : p, q \in \mathbb{R} \right\}.$$

Note in this case, the solution set U can be written in terms of linear combinations of the vectors $(2,1,0,0,)^T$ and $(-1,0,2,1)^T$, that is:

$$U = \left\{ p \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + q \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : p, q \in \mathbb{R} \right\} = \text{Span} \left(\left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \right\} \right).$$

There are several lessons that we can learn from comparing these four systems of linear equations.

I) The last system was the simplest to solve because all the terms on the RHS are zero. We thus will make a distinction between systems.

Definition 1: Homogeneous and inhomogeneous systems

We say that a system of linear equations $(*)$ is **homogeneous** to mean that all the terms on the RHS of the equations in $(*)$ are zero, that is $\mathbf{b} = \mathbf{0}$.

We say that a system of linear equations $(*)$ is **inhomogeneous** to mean that the terms on the RHS of the equations in $(*)$ are not all equal to zero, i.e. $\mathbf{b} \neq \mathbf{0}$.
Some authors use the terminology **non-homogeneous** for inhomogeneous.

Definition 2: Nullspace

Let A be the coefficient matrix of a homogeneous system of linear equations and let S be its solution set. The **nullspace** of A, denoted, $N(A)$, means the solution set S.

The next two comments are about the solution sets, and so they are reproduced here:

$$S = \left\{ \begin{pmatrix} 2+2s-t \\ s \\ 1+2t \\ t \end{pmatrix} : \ s, t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : \ s, t \in \mathbb{R} \right\}.$$

$$T = \left\{ \begin{pmatrix} 1+2a-b \\ a \\ 2+2b \\ b \end{pmatrix} : \ a, b \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + a \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : \ a, b \in \mathbb{R} \right\}.$$

$$U=\left\{\begin{pmatrix}2p-q\\p\\2q\\q\end{pmatrix}:\ p,q\in\mathbb{R}\right\}=\left\{\begin{pmatrix}0\\0\\0\\0\end{pmatrix}+p\begin{pmatrix}2\\1\\0\\0\end{pmatrix}+q\begin{pmatrix}-1\\0\\2\\1\end{pmatrix}:\ p,q\in\mathbb{R}\right\}$$

$$=\left\{p\begin{pmatrix}2\\1\\0\\0\end{pmatrix}+q\begin{pmatrix}-1\\0\\2\\1\end{pmatrix}:p,q\in\mathbb{R}\right\}=Span\left(\left\{\begin{pmatrix}2\\1\\0\\0\end{pmatrix},\begin{pmatrix}-1\\0\\2\\1\end{pmatrix}\right\}\right).$$

Note that $U=N(A)$, where $A=\begin{pmatrix}1&-2&-1&3\\2&-4&1&0\\1&-2&2&-3\end{pmatrix}$.

II) We notice that the solution set for the homogeneous system of equations can be written as the Span of a set of vectors. The solution sets for the two inhomogeneous systems of equations cannot be written as the Span of a set of vectors.

III) We also notice that the solution sets S,T and U, are very similar. To be more precise: the solution sets of the inhomogeneous systems, S and T, differ by only a constant vector, and the part which they have in common is U.
We will make these observations more precise in the next lecture, and state them as a theorem.

IV) The systems in all four examples have the same coefficient matrix. There is a practical way of dealing with more than one system with the same coefficient matrix.

Suppose we did actually have to solve the four systems in the previous 4 examples, simultaneously. We have seen that the elementary row operations required to obtain the row echelon form (and the RREF) are the same in all cases, so why not do it all in one go?

Given

$$\begin{cases}x-2y-z+3w=1\\2x-4y+z=5\\x-2y+2z-3w=4\end{cases}\text{ with augmented matrix }\left(\begin{array}{cccc|c}1&-2&-1&3&1\\2&-4&1&0&5\\1&-2&2&-3&4\end{array}\right),$$

and

$$\begin{cases}x-2y-z+3w=1\\2x-4y+z=2\\x-2y+2z-3w=3\end{cases}\text{ with augmented matrix }\left(\begin{array}{cccc|c}1&-2&-1&3&1\\2&-4&1&0&2\\1&-2&2&-3&3\end{array}\right),$$

and

$$\begin{cases}x-2y-z+3w=-1\\2x-4y+z=4\\x-2y+2z-3w=5\end{cases}\text{ with augmented matrix }\left(\begin{array}{cccc|c}1&-2&-1&3&-1\\2&-4&1&0&4\\1&-2&2&-3&5\end{array}\right),$$

and

$$\left\{\begin{aligned} x-2y-z+3w&=0\\ 2x-4y+z&=0\\ x-2y+2z-3w&=0 \end{aligned}\right. \quad \text{with augmented matrix} \quad \left(\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & 0\\ 2 & -4 & 1 & 0 & 0\\ 1 & -2 & 2 & -3 & 0 \end{array}\right).$$

We can consider the "super-augmented matrix"

$$\left(\begin{array}{rrrr|rrrr} 1 & -2 & -1 & 3 & 1 & 1 & -1 & 0\\ 2 & -4 & 1 & 0 & 5 & 2 & 4 & 0\\ 1 & -2 & 2 & -3 & 4 & 3 & 5 & 0 \end{array}\right).$$

This super-augmented matrix contains the information content of ALL these systems in the same place. We can row reduce this super-augmented matrix to row echelon form or to reduced row echelon form, and then we can CAREFULLY peel off the information about each system separately.

First we will perform Gaussian elimination.

$$\left\{\begin{array}{l} R_2 \to R_2 - 2R_1\\ R_3 \to R_3 - R_1 \end{array}\right. \quad \text{give} \quad \left(\begin{array}{rrrr|rrrr} 1 & -2 & -1 & 3 & 1 & 1 & -1 & 0\\ 0 & 0 & 3 & -6 & 3 & 0 & 6 & 0\\ 0 & 0 & 3 & -6 & 3 & 2 & 6 & 0 \end{array}\right).$$

$$R_2 \to \tfrac{1}{3}R_2 \quad \text{gives} \quad \left(\begin{array}{rrrr|rrrr} 1 & -2 & -1 & 3 & 1 & 1 & -1 & 0\\ 0 & 0 & 1 & -2 & 1 & 0 & 2 & 0\\ 0 & 0 & 3 & -6 & 3 & 2 & 6 & 0 \end{array}\right).$$

$$R_3 \to R_3 - 3R_2 \quad \text{yields the REF} \quad \left(\begin{array}{rrrr|rrrr} 1 & -2 & -1 & 3 & 1 & 1 & -1 & 0\\ 0 & 0 & 1 & -2 & 1 & 0 & 2 & 0\\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \end{array}\right).$$

$$R_1 \to R_1 + R_2 \quad \text{gives the RREF} \quad \left(\begin{array}{rrrr|rrrr} 1 & -2 & 0 & 1 & 2 & 1 & 1 & 0\\ 0 & 0 & 1 & -2 & 1 & 0 & 2 & 0\\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \end{array}\right).$$

From the reduced row echelon form of the "super-augmented matrix", we can now peel off the four equivalent systems.

For the **first** system we need the reduced coefficient matrix and the **first** column after the dashed vertical line.

$$\left(\begin{array}{rrrr|r} 1 & -2 & 0 & 1 & 2\\ 0 & 0 & 1 & -2 & 1\\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This corresponds to the system of linear equations:

$$\begin{aligned} x - 2y + w &= 2 \\ z - 2w &= 1 \\ 0 &= 0 \end{aligned}$$

with solution set

$$S = \left\{ \begin{pmatrix} 2+2s-t \\ s \\ 1+2t \\ t \end{pmatrix} \ : \ s,t \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s,t \in \mathbb{R} \right\}.$$

For the **second** system, we need the reduced coefficient matrix and the **second** column after the dashed vertical line.

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

This corresponds to the system of linear equations:

$$\begin{aligned} x - 2y + w &= 1 \\ z - 2w &= 0 \\ 0 &= 2 \end{aligned}$$

which is inconsistent and has solution set of $\emptyset$.

For the **third** system we need the reduced coefficient matrix and the **third** column after the dashed vertical line,

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

This corresponds to the system of equations:

$$\begin{aligned} x - 2y + w &= 1 \\ z - 2w &= 2 \\ 0 &= 0 \end{aligned}$$

with solution set

$$T = \left\{ \begin{pmatrix} 1+2a-b \\ a \\ 2+2b \\ b \end{pmatrix} \;:\; a, b \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + a \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \;:\; a, b \in \mathbb{R} \right\}.$$

For the **fourth** system we need the reduced coefficient matrix and the **fourth** column after the dashed vertical line,

$$\left(\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right).$$

This corresponds to the system of linear equations:

$$\begin{aligned} x - 2y + w &= 0 \\ z - 2w &= 0 \\ 0 &= 0 \end{aligned}$$

with solution set

$$\begin{aligned} U = \left\{ \begin{pmatrix} 2p-q \\ p \\ 2q \\ q \end{pmatrix} \;:\; p, q \in \mathbb{R} \right\} &= \left\{ p \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + q \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \;:\; p, q \in \mathbb{R} \right\} \\ &= \operatorname{Span} \left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \right\}. \end{aligned}$$

Topic 10B

Example Using Complex Numbers

We consider the following system of two linear equations in two complex unknowns with complex coefficients, and we perform Gauss-Jordan elimination on it. The only new feature is that the arithmetic is much more intricate, because of the complex numbers, and so there is much more chance of making a simple arithmetical error.

Example 1

$$\begin{cases} (1+i)\,z - (2+3i)w = -15i & (e_1) \\ (4+5i)\,z + (3-2i)w = 37-7i & (e_2) \end{cases}$$

The augmented matrix is:

$$\left(\begin{array}{cc|c} 1+i & -2-3i & -15i \\ 4+5i & 3-2i & 37-7i \end{array} \right).$$

We see at once that z is a basic variable and that the $(1,1)$ position is a pivot position. We scale the pivot to unity, i.e. commence with the elementary row operation:

$$R_1 \to \frac{1}{1+i} R_1 \quad \left(= \left(\frac{1}{2} - \frac{i}{2} \right) R_1 \right) \quad \text{to get} \quad \left(\begin{array}{cc|c} 1 & \frac{-5}{2} - \frac{i}{2} & \frac{-15}{2} - \frac{15i}{2} \\ 4+5i & 3-2i & 37-7i \end{array} \right).$$

We now have a leading 1 as our pivot in the $(1,1)$ position. We then perform

$$R_2 \to R_2 - (4+5i)R_1 \quad \text{to get} \quad \left(\begin{array}{cc|c} 1 & \frac{-5}{2} - \frac{i}{2} & \frac{-15}{2} - \frac{15i}{2} \\ 0 & \frac{21}{2} + \frac{25i}{2} & \frac{59}{2} + \frac{121i}{2} \end{array} \right).$$

This matrix is in row echelon form. The ranks of both the coefficient matrix and the augmented matrix are 2, and so the system is consistent. There will be $2-2=0$ parameters in the solution set. You could stop here, although it makes sense to at least do the division by $\left(\frac{21}{2} + \frac{25i}{2}\right)$ step, and convert the second pivot to unity:

$$R_2 \to \frac{2}{21+25i} R_2 \quad \left[= \left(\frac{21}{533} - \frac{25i}{533} \right) R_2 \right] \quad \text{to get} \quad \left(\begin{array}{cc|c} 1 & \frac{-5}{2} - \frac{i}{2} & \frac{-15}{2} + \frac{-15i}{2} \\ 0 & 1 & 4+i \end{array} \right).$$

This matrix is also in row echelon form.

The last step of the Gauss-Jordan algorithm is

$$R_1 \to R_1 + \left(\frac{5}{2} + \frac{i}{2}\right) R_2 \quad \text{to get} \quad \left(\begin{array}{cc|c} 1 & 0 & 2-3i \\ 0 & 1 & 4+i \end{array}\right).$$

This matrix is in reduced row echelon form.

We conclude that $z = 2 - 3i$ and $w = 4 + i$, and the solution set S is

$$S = \left\{ \begin{pmatrix} 2-3i \\ 4+i \end{pmatrix} \right\}.$$

We conclude by checking that we really do have a solution by substituting z and w back in the original system of equations.

$$(1+i)\,(2-3i) - (2+3i)(4+i) = -15i$$

and so (e_1) holds true, and also

$$(4+5i)\,(2-3i) + (3-2i)(4+i) = 37 - 7i.$$

and so (e_2) holds true.

Topic 11A

Matrix Multiplication I

Multiplying a Vector by a Matrix

Decomposing a matrix.

Let $A \in M_{m\times n}$. This means that A is a matrix with m rows and n columns.
With matrices we always refer to rows first and then columns.

One way to think of the matrix is that it is an array constructed from its individual entries. The $(i,j)^{th}$ **entry** of the matrix A is denoted either by $(A)_{ij}$ or by a_{ij} and means the entry in the i^{th} row and j^{th} column of the matrix A. We often write the matrix as

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \\ \vdots & \vdots & \ddots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Definition 1: Row vector

Let $G \in M_{1\times n}$. We will then call the matrix G, a **row vector**.

We will also refer to this vector by $\mathbf{G}$, i.e., we will bold it (or underline when writing on paper), so that we think of vectors.

We will use a capital letter to distinguish it from a column vector, and also keep the same letter of the alphabet.

We will write the j^{th} entry in $\mathbf{G}$ as $\mathbf{G}_j$ so this is the entry in the j^{th} column of the row vector $\mathbf{G}$.

Note: if $G \in M_{1\times n}$, then $G = \mathbf{G}$ and $(G)_{1j} = \mathbf{G}_j$.

Example 1

Let $C = \mathbf{C} = (5,\ -2, 0)$ be a **row vector**.

Then $(C)_{11} = \mathbf{C}_1 = c_{11} = 5$. In addition, $\mathbf{C}_2 = -2$ and $\mathbf{C}_3 = 0$.

Example 2

Let $D = \begin{pmatrix} 1 & 2 \\ -3 & -4 \\ 7 & 9 \end{pmatrix}$ be a (3×2) matrix.

Then, $d_{12} = (D)_{12} = 2$, $d_{22} = (D)_{22} = -4$, and $d_{31} = (D)_{31} = 7$.

Notation 1: Rows of a matrix.

A second way to decompose a matrix is to imagine it to be constructed from its rows.
We will use the notation $\mathbf{A}^i$ to indicate the i^{th} row of a matrix $A \in M_{m\times n}(\mathbb{F})$, where $m > 1$.

It follows that $\mathbf{A}^i$ is a row vector with n entries, that is to say that $\mathbf{A}^i \in M_{1\times n}$, where $i = 1, \cdots, m$.

$$\begin{aligned} \mathbf{A}^1 &= \begin{pmatrix} a_{11}\ , & a_{12}\ , & \ldots\ , & a_{1n} \end{pmatrix} \\ \mathbf{A}^2 &= \begin{pmatrix} a_{21}\ , & a_{22}\ , & \ldots\ , & a_{2n} \end{pmatrix} \\ &\vdots \\ \mathbf{A}^m &= \begin{pmatrix} a_{m1}\ , & a_{m2}\ , & \ldots\ , & a_{mn} \end{pmatrix} \end{aligned}$$

We can then write:

$$A = \begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^m \end{pmatrix}.$$

Example 3

Let $D = \begin{pmatrix} 1 & 2 \\ -3 & -4 \\ 7 & 9 \end{pmatrix}$,

Then, $\mathbf{D}^1 = \begin{pmatrix} 1\ , & 2 \end{pmatrix}$, $\mathbf{D}^2 = \begin{pmatrix} -3\ , & -4 \end{pmatrix}$, $\mathbf{D}^3 = \begin{pmatrix} 7\ , & 9 \end{pmatrix}$, and $D = \begin{pmatrix} \mathbf{D}^1 \\ \mathbf{D}^2 \\ \mathbf{D}^3 \end{pmatrix}$.

Notation 2: Columns of a matrix.

A third way that we can think of the matrix $A \in M_{m\times n}(\mathbb{F})$, is that it is an array which is constructed from n columns, denoted by $\mathbf{a}_1\ ,\ \mathbf{a}_2\ ,\ \cdots\ ,\ \mathbf{a}_n$, with $\mathbf{a}_j \in M_{m\times 1}(= \mathbb{F}^m)$.

We then have:

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \ \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \cdots, \ \mathbf{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix},$$

and we can write:

$$A = (\mathbf{a}_1\,,\ \mathbf{a}_2\,,\ \cdots\,,\ \mathbf{a}_n)\,.$$

Example 4

Let $D = \begin{pmatrix} 1 & 2 \\ -3 & -4 \\ 7 & 9 \end{pmatrix}$, then $\mathbf{d}_1 = \begin{pmatrix} 1 \\ -3 \\ 7 \end{pmatrix}$ and $\mathbf{d}_2 = \begin{pmatrix} 2 \\ -4 \\ 9 \end{pmatrix}$, $D = (\mathbf{d}_1, \mathbf{d}_2)$.

There are many other ways in which we can decompose or partition a matrix, depending upon its actual size. For example, it might be convenient to think of a (4×4) matrix to be constructed from the 4, (2×2) matrices, one in each corner.

We will now define the multiplication of a (column) vector in $\mathbb{F}^n$ by a $(m \times n)$ matrix.

Definition 2a: Matrix multiplication in terms of the individual entries.

Let $A \in M_{m\times n}$ and $\mathbf{x} \in \mathbb{F}^n$. We then define the product $A\mathbf{x}$ as follows:

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

The expression appearing on the right hand side here is an old friend. It is just the LHS of our standard system of linear equations, called system $(*)$ in previous lectures.

Notice that the sizes of the matrix and the vector must be compatible otherwise this just does not work. That is, the matrix A has n columns and the vector $\mathbf{x}$ has n entries.

Definition 2b: Matrix multiplication in terms of columns.

Notice that we can also write matrix multiplication as:

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (\mathbf{a}_1,\ \mathbf{a}_2, \cdots,\ \mathbf{a}_n)\,\mathbf{x}$$

$$= x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n,$$

that is, the product of the matrix A and the column vector $\mathbf{x}$ yields a **linear combination** of the columns (vectors) of the matrix A, with the entries in the vector $\mathbf{x}$ determining the scalar multiples.

Definition 2c: Matrix multiplication in terms of rows.

Notice that we can also write matrix multiplication as:

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^m \end{pmatrix} \mathbf{x}$$

$$= \begin{pmatrix} \mathbf{A}^1_{\ 1}\, x_1 + \mathbf{A}^1_{\ 2}\, x_2 + \cdots + \mathbf{A}^1_{\ n}\, x_n \\ \mathbf{A}^2_{\ 1}\, x_1 + \mathbf{A}^2_{\ 2}\, x_2 + \cdots + \mathbf{A}^2_{\ n}\, x_n \\ \vdots \\ \mathbf{A}^m_{\ 1}\, x_1 + \mathbf{A}^m_{\ 2}\, x_2 + \cdots + \mathbf{A}^m_{\ n}\, x_n \end{pmatrix},$$

that is, the product of the matrix $A \in M_{m\times n}$ and the column vector $\mathbf{x} \in M_{n\times 1}$ is a column vector in $\mathbb{F}^m$ and a matrix in $M_{m\times 1}$.

The entry in the i^{th} row is obtained by multiplying the entries in the i^{th} row of the matrix A with the corresponding entries of the column vector $\mathbf{x}$.

$$(A\mathbf{x})_i = a_{i1}\, x_1 + a_{i2}\, x_2 + \cdots + a_{in}\, x_n = \sum_{j=1}^{n} a_{ij}\, x_j = \sum_{j=1}^{n} \mathbf{A}^i_{\ j}\, x_j.$$

Practically this is how we usually do matrix multiplication: "multiply" rows by columns.

Examples 5

We multiply $\begin{pmatrix} 1 \\ -4 \\ 6 \end{pmatrix}$ by $\begin{pmatrix} 1 & 6 & 1 \\ 3 & 4 & 5 \\ 5 & 2 & -3 \end{pmatrix}$ in two different ways:

$$\begin{pmatrix} 1 & 6 & 1 \\ 3 & 4 & 5 \\ 5 & 2 & -3 \end{pmatrix}\begin{pmatrix} 1 \\ -4 \\ 6 \end{pmatrix} = (1)\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + (-4)\begin{pmatrix} 6 \\ 4 \\ 2 \end{pmatrix} + (6)\begin{pmatrix} 1 \\ 5 \\ -3 \end{pmatrix} = \begin{pmatrix} -17 \\ 17 \\ -21 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 6 & 1 \\ 3 & 4 & 5 \\ 5 & 2 & -3 \end{pmatrix}\begin{pmatrix} 1 \\ -4 \\ 6 \end{pmatrix} = \begin{pmatrix} (1)(1) + (6)(-4) + (1)(6) \\ (3)(1) + (4)(-4) + (5)(6) \\ (5)(1) + (2)(-4) + (-3)(6) \end{pmatrix} = \begin{pmatrix} -17 \\ 17 \\ -21 \end{pmatrix},$$

Examples 6

We multiply $\begin{pmatrix} 1 \\ 1-i \\ 2-3i \end{pmatrix}$ by $\begin{pmatrix} 1+i & 2+2i & 3-i \\ 2+3i & 4+i & 5-2i \end{pmatrix}$ in two different ways:

$$\begin{aligned} \begin{pmatrix} 1+i & 2+2i & 3-i \\ 2+3i & 4+i & 5-2i \end{pmatrix}\begin{pmatrix} 1 \\ 1-i \\ 2-3i \end{pmatrix} &= (1)\begin{pmatrix} 1+i \\ 2+3i \end{pmatrix} + (1-i)\begin{pmatrix} 2+2i \\ 4+i \end{pmatrix} + \\ &\quad (2-3i)\begin{pmatrix} 3-i \\ 5-2i \end{pmatrix} \\ &= \begin{pmatrix} 1+i \\ 2+3i \end{pmatrix} + \begin{pmatrix} 4 \\ 5-3i \end{pmatrix} + \begin{pmatrix} 3-11i \\ 4-19i \end{pmatrix} \\ &= \begin{pmatrix} 8-10i \\ 11-19i \end{pmatrix}. \end{aligned}$$

$$\begin{aligned} \begin{pmatrix} 1+i & 2+2i & 3-i \\ 2+3i & 4+i & 5-2i \end{pmatrix}\begin{pmatrix} 1 \\ 1-i \\ 2-3i \end{pmatrix} &= \begin{pmatrix} (1+i)(1) + (2+2i)(1-i) + (3-i)(2-3i) \\ (2+3i)(1) + (4+i)(1-i) + (5-2i)(2-3i) \end{pmatrix} \\ &= \begin{pmatrix} 8-10i \\ 11-19i \end{pmatrix}. \end{aligned}$$

Lemma 1: Linearity of matrix multiplication

Let $A \in M_{m\times n}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, $c \in \mathbb{F}$. Then:

$$\begin{cases} A\,(\mathbf{x}+\mathbf{y}) = A\,\mathbf{x} + A\,\mathbf{y} \\ A\,(c\,\mathbf{x}) = c\,A\,\mathbf{x} \end{cases}$$

The proof of these two statements follows immediately from the definition of matrix multiplication; properties of multiplication of scalars, and the definition of addition of vectors in $\mathbb{F}^n$.

These two properties of matrix multiplication are very natural.

Remarks 1 on Matrix Multiplication

The components of the vector which results under matrix multiplication can be expressed in terms of the dot product when you are working in $\mathbb{R}^n$, or the standard inner product when working in $\mathbb{C}^n$.

We will not be making use of the following expressions in this course, however they will be used in the construction of some of the proofs in the next course: MATH235.

We have that if $A \in M_{m\times n}(\mathbb{R})$ and $\mathbf{x} \in \mathbb{R}^n$, then

$$(A\mathbf{x})_i = \left(\mathbf{A}^i\right)^T \bullet \mathbf{x}.$$

We have that if $C \in M_{m\times n}(\mathbb{C})$ and $\mathbf{z} \in \mathbb{C}^n$, then

$$(C\mathbf{z})_i = \left\langle \left(\mathbf{C}^i\right)^T, \bar{\mathbf{z}} \right\rangle.$$

We can combine these two expressions in one equation:

if $D \in M_{m\times n}$ and $\mathbf{w} \in \mathbb{C}^n$ or $\mathbb{R}^n$, then

$$(D\mathbf{w})_i = \left\langle \left(\mathbf{D}^i\right)^T, \bar{\mathbf{w}} \right\rangle.$$

Topic 11B
Solution Sets

Consider a system of linear equations: the system $(*) : A\mathbf{x} = \mathbf{b}$. We will be looking at this system when $\mathbf{b} = \mathbf{0}$ and will examine the relationship between the solution set when $\mathbf{b} = \mathbf{0}$ and the solution set when $\mathbf{b} \neq \mathbf{0}$.

Definition 3: Homogeneous

We say that the system of linear equations, $A\mathbf{x} = \mathbf{b}$, is **homogeneous** to mean that $\mathbf{b} = \mathbf{0}$.

Note that since $A\mathbf{0} = \mathbf{0}$, the solution set to a homogeneous system of linear equations always contains the zero vector, and it is never empty.
Thus, **a homogeneous system of linear equations is always consistent**.
The solution $\mathbf{x} = \mathbf{0}$ is called the **trivial solution**.

Definition 4: Inhomogeneous

We say that the system of linear equations, $A\mathbf{x} = \mathbf{b}$, is **inhomogeneous** or **non-homogeneous** , to mean that $\mathbf{b} \neq \mathbf{0}$.
The solution set of an inhomogeneous system does not contain the trivial solution.
Inhomogeneous systems of linear equations may be consistent or inconsistent.

Definition 5: Associated homogeneous system

Suppose you have an inhomogeneous system of linear equations, that is $C\mathbf{x} = \mathbf{d}$, with $\mathbf{d} \neq \mathbf{0}$.

We refer to the linear system $C\mathbf{x} = \mathbf{0}$ as the **associated homogeneous (linear) system** (to $C\mathbf{x} = \mathbf{d}$).

Examples 7

$\begin{pmatrix} 1 & 2 & 3 \\ -4 & -5 & 7 \\ 2 & -4 & 6 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ is a homogeneous system of 3 linear equations in 3 unknowns.

$\begin{pmatrix} 1 & 2 & 3 & 4 \\ -4 & -5 & 7 & -9 \\ 2 & -4 & 6 & -4 \end{pmatrix} \mathbf{y} = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$ is an inhomogeneous system of 3 linear equations in 4 unknowns.

$\begin{pmatrix} 1 & 2 \\ -4 & -5 \end{pmatrix} \mathbf{z} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is the homogeneous system of equations associated to the inhomogeneous system of linear equations, $\begin{pmatrix} 1 & 2 \\ -4 & -5 \end{pmatrix} \mathbf{z} = \begin{pmatrix} 1 \\ -6 \end{pmatrix}$.

As we have seen through our examples, the solution sets of an inhomogeneous system of linear equations, and that of its associated homogeneous system are closely related.
We will make this explicit through a number of lemmas.

Suppose we have an inhomogeneous system $A\mathbf{x} = \mathbf{b}$, with $\mathbf{b} \neq \mathbf{0}$, with a solution set $\tilde{S}$, and its associated homogeneous linear system, $A\mathbf{x} = \mathbf{0}$, with solution set S.

Lemma 2:

Linear combinations of solutions to a homogeneous system of linear equations are also solutions to the homogeneous system, that is,

If $\mathbf{x}_1$ and $\mathbf{x}_2 \in S$, and if $a_1 \in \mathbb{F}$, then $(\mathbf{x}_1 + \mathbf{x}_2) \in S$ and $a_1\mathbf{x}_1 \in S$.

Proof

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be solution to a homogeneous system, i.e. $\mathbf{x}_1$, $\mathbf{x}_2 \in S$.
We thus have: $A\mathbf{x}_1 = \mathbf{0}$ and $A\mathbf{x}_2 = \mathbf{0}$.

Since $A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2$, and $A(a_1\mathbf{x}_1) = a_1\, A\mathbf{x}_1$ using Lemma 1 (Topic 11A), we get:

$$A(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0} \text{ and } A(a_1\mathbf{x_1}) = a_1(\mathbf{0}) = \mathbf{0}.$$

And thus we conclude that both $(\mathbf{x}_1 + \mathbf{x}_2) \in S$ and $a_1\mathbf{x}_1 \in S$. ■

Note that we can combine these two results in Lemma 2 and state:

$$\text{If } \mathbf{x}_1, \mathbf{x}_2 \in S \text{ and if } a_1,\, a_2 \in \mathbb{F}, \text{ then } (a_1\mathbf{x}_1 + a_2\mathbf{x}_2) \in S.$$

We can also extend this result inductively to n vectors in S and n scalars in $\mathbb{F}$. In other words, if you have n solutions of a homogeneous system of linear equations, then you may form any linear combinations of these solution vectors to produce another solution of the homogeneous system of linear equations.

Lemma 3: Relation between $\tilde{S}$ and S part I.

If we have two solutions of an inhomogeneous system of linear equations, then their difference is a solution to the associated homogeneous system of these linear equations, that is,

If $\mathbf{y}_1$ and $\mathbf{y}_2 \in \tilde{S}$, then $(\,\mathbf{y}_1 - \mathbf{y}_2\,) \in S$.

Proof

Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be solutions to an inhomogeneous system , i.e. $\mathbf{y_1}\,,\,\mathbf{y_2} \in \tilde{S}$.

We thus have: $\quad A\mathbf{y}_1 = \mathbf{b}$ and $\;A\mathbf{y}_2 = \mathbf{b}$.

Since $A(\mathbf{y}_1 - \mathbf{y}_2) = A\mathbf{y}_1 - A\mathbf{y}_2$, using Lemma 1 (Topic 11A), it follows that

$$A(\mathbf{y}_1 - \mathbf{y}_2) = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

And thus we conclude that $(\mathbf{y}_1 - \mathbf{y}_2) \in S$. ■

In other words, Lemma 3 guarantees that if you already have two solutions of an inhomogeneous system of linear equations, then their difference is a solution of its associated homogeneous system.

Definition 6: Particular solution

Suppose we have a consistent system of linear equations, $A\mathbf{x} = \mathbf{b}$.

We refer to a solution $\mathbf{x}_p$ of this system, i.e. $A\mathbf{x}_p = \mathbf{b}$, as a **particular solution** to this system.

Note that there may be many particular solutions to the consistent system $A\mathbf{x} = \mathbf{b}$.

Lemma 4: Relation between $\tilde{S}$ and S part II.

The solution set $\tilde{S}$ can be constructed from the solution set S and a single particular solution, that is,

If $\mathbf{y}_p \in \tilde{S}$, then $\tilde{S} = \{\mathbf{y}_p + \mathbf{x} : \mathbf{x} \in S\}$.

Proof:

Let $\mathbf{y}_p$ be a particular solution to the inhomogeneous system, i.e. $A\mathbf{y}_p = \mathbf{b}$.

Let $\mathbf{x}$ be a solution of its associated homogeneous system, i.e. $A\mathbf{x} = \mathbf{0}$.
We have to show that two sets are equal. We do this in the standard way of showing that each set is contained in the other.

a) Suppose that $\mathbf{y} \in \tilde{S}$. Then we know that $A\mathbf{y} = \mathbf{b}$.

It follows from Lemma 3 that $(\mathbf{y} - \mathbf{y}_p) \in S$,
that is, $\mathbf{y} - \mathbf{y}_p = \mathbf{x}_1$, for some $\mathbf{x}_1 \in S$, and therefore, $\mathbf{y} = \mathbf{y}_p + \mathbf{x}_1$, where $\mathbf{x}_1 \in S$.
We then conclude that $\mathbf{y} \in \{\mathbf{y}_p + \mathbf{x} : \mathbf{x} \in S\}$.

b) Suppose now that $\mathbf{z} \in \{\mathbf{y}_p + \mathbf{x} : \mathbf{x} \in S\}$. We can then write $\mathbf{z} = \mathbf{y}_p + \mathbf{x}_2$, for some $\mathbf{x}_2 \in S$.

Using Lemma 1 (Topic 11A), we have that $A\mathbf{z} = A(\mathbf{y}_p + \mathbf{x}_2) = A\mathbf{y}_p + A\mathbf{x}_2$, that is,
$A\mathbf{z} = \mathbf{b} + \mathbf{0} = \mathbf{b}$. We thus conclude that $\mathbf{z} \in \tilde{S}$.
Therefore, the two sets $\tilde{S}$ and $\{\mathbf{y}_p + \mathbf{x} : \mathbf{x} \in S\}$ are identical. $\blacksquare$

In other words, Lemma 4 guarantees that if you have a single particular solution to the inhomogeneous system $A\mathbf{x} = \mathbf{b}$, and you also have the solution set to its associated homogeneous system, $A\mathbf{x} = \mathbf{0}$, then you can write the complete solution set $\tilde{S}$ to $A\mathbf{x} = \mathbf{b}$.

Example 8

Let us reconsider Examples 3 and 4 in Topic 10A.
They were 2 different systems of linear equations, with the same coefficient matrix, A:

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

The solution set to the homogeneous system $A\mathbf{x} = \mathbf{0}$ was the set U, (in Ex. 4):

$$U = \left\{ s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : \ s, t \in \mathbb{R} \right\}.$$

The solution set to the inhomogeneous system $A\mathbf{x} = (-1, 4, 5)^T$ was the set T, (in Ex. 3):

$$T = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : \ s, t \in \mathbb{R} \right\}.$$

We first illustrate Lemma 2.

By choosing particular values of the parameters, we get two solutions of the homogeneous system, that is: $\mathbf{x}_1 = (2,1,0,0)^T \in U$ and $\mathbf{x}_2 = (-1,0,2,1)^T \in U$.

If we now let $\mathbf{z} = 2\mathbf{x}_1 + 3\mathbf{x}_2 = (1,2,6,3)^T$, then $A\mathbf{z} = \mathbf{0}$.

And we conclude, as expected by Lemma 2, that $\mathbf{z} \in U$.

Secondly, we illustrate Lemma 3.

Let us refer to the system $A\mathbf{x} = (\,-1,4,5)^T$, with solution set T.

By choosing $s = t = 0$ and then $s = t = 1$, we get two particular solutions:

$$\mathbf{y}_1 = (1,0,2,0)^T \quad \text{and} \quad \mathbf{y}_2 = (2,1,4,1)^T$$

.

We notice that $A\,(\mathbf{y}_1 - \mathbf{y}_2) = A\ (-1,-1,-2,-1)^T = \mathbf{0}$.

And we conclude, as expected by Lemma 3, that $(\mathbf{y}_1 - \mathbf{y}_2) \in U$.

Finally, we illustrate Lemma 4.

The vector $\mathbf{y}_p = (1,0,2,0)^T$ is a particular solution of the system $A\mathbf{x} = (-1,4,5)^T$.

Note that we already have the solution set to its associated homogeneous system

$$U = \left\{ s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s,t \in \mathbb{R} \right\}.$$

We conclude that we can write the solution set T of the system $A\mathbf{x} = (-1,4,5)^T$ as:

$$T = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s,t \in \mathbb{R} \right\}.$$

We notice from our examples, that if we have already obtained the solution set $\tilde{S}$ of an inhomogeneous system, then we can obtain the solution S of its associated homogeneous system, by subtracting $\mathbf{y}_p$ from all the solutions in $\tilde{S}$, that is:

$$\text{If } \mathbf{y}_p \in \tilde{S}, \ \text{then } S = \left\{ \mathbf{z} - \mathbf{y}_p : \mathbf{z} \in \tilde{S} \right\}.$$

Example 9

Let us continue Example 8 by illustrating the above statement on how to obtain the solution set of a homogeneous system given a particular solution to the inhomogeneous system.

The solution set to the inhomogeneous system $A\mathbf{x} = (-1, 4, 5)^T$ is the set T, with

$$T = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\}.$$

We know that $\mathbf{y}_1 = (1, 0, 2, 0)^T$ is a particular solution to $A\mathbf{x} = (-1, 4, 5)^T$, thus we can immediately write down the solution set to the associated homogeneous equation $A\mathbf{x} = \mathbf{0}$:

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\}$$

$$= \left\{ s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\}.$$

This is of course the set U given in Example 8.

Note that we could have chosen another particular solution, say $\mathbf{y}_2 = (2, 1, 4, 1)^T$. The solution set to the associated homogeneous equation would then be:

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ 4 \\ 1 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\}$$

$$= \left\{ \begin{pmatrix} -1 \\ -1 \\ -2 \\ -1 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\}.$$

But this can be written as

$$\left\{ (s-1) \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + (t-1) \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\} = \left\{ p \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + q \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : p, q \in \mathbb{R} \right\},$$

where we have let $p = s-1$ and $q = t-1$. We immediately recognize this set to be the set U.

We conclude by relating the solution sets of two different systems of equations with the same coefficient matrices but having different right-hand sides.

Lemma 5: Relation between the solution sets of two different inhomogeneous systems with the same coefficient matrix.

Consider the two different inhomogeneous systems

$$A\mathbf{x} = \mathbf{b}, \quad \text{and } A\mathbf{x} = \mathbf{c},$$

with solution sets $\tilde{S}_1$ and $\tilde{S}_2$, respectively, and with particular solutions $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively. We then have:

$$\tilde{S}_2 = \left\{ \mathbf{p}_2 + (\mathbf{z} - \mathbf{p}_1) : \mathbf{z} \in \tilde{S}_1 \right\}.$$

In practice, we can usually go from $\tilde{S}_1$ to $\tilde{S}_2$ by inspection, especially if $\mathbf{p}_2$ is known: that is,

$$\begin{aligned} \text{If } \tilde{S}_1 &= \{\mathbf{p}_1 + a_1\mathbf{w}_1 + \cdots + a_q\mathbf{w}_q : \ a_1, a_2, \ldots, a_q \in \mathbb{F}\}, \\ \text{then } \tilde{S}_2 &= \{\mathbf{p}_2 + a_1\mathbf{w}_1 + \cdots + a_q\mathbf{w}_q : \ a_1, a_2, \ldots, a_q \in \mathbb{F}\}. \end{aligned}$$

Example 10

Let us extend Example 8 and consider the coefficient matrix:

$$A = \begin{pmatrix} 1 & -2 & -1 & 3 \\ 2 & -4 & 1 & 0 \\ 1 & -2 & 2 & -3 \end{pmatrix}$$

The solution set to the inhomogeneous system $A\mathbf{x} = (-1, 4, 5)^T$ is the set T, with

$$T = \left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\},$$

and the solution set to the inhomogeneous system $A\mathbf{x} = (1, 5, 4)^T$ is the set S, with

$$S = \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \ : \ s, t \in \mathbb{R} \right\}.$$

The vector $(2, 0, 1, 0)^T$ is a particular solution to the system $A\mathbf{x} = (1, 5, 4)^T$.

The vector $(1,0,2,0)^T$ is a particular solution to the system $A\mathbf{x} = (-1,4,5)^T$.

Once we know T, then we can obtain S by simply replacing the vector $(1,0,2,0)^T$ by the vector $(2,0,1,0)^T$.

Topic 11C

Matrix Multiplication II

Multiplication of Matrices

We have already examined the process of multiplying a column vector, $\mathbf{x}$, by a matrix, A, as long as their sizes are compatible, i.e. we need $\mathbf{x} \in \mathbb{F}^n$ and $A \in M_{m\times n}(\mathbb{F})$. We can extend this process to multiplying several column vectors, of the correct size, by the same matrix simultaneously. This is exactly what happens when we perform matrix multiplication.

Definition 7: Matrix multiplication.

Let $A \in M_{m\times n}$ and $B \in M_{n\times p}$. We then define the product $AB = C$.

This new matrix, $C \in M_{m\times p}$, and is constructed as follows:

$$C = AB = A\,(\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_p) = (A\mathbf{b}_1, A\mathbf{b}_2, \cdots, A\mathbf{b}_p).$$

That is, the j^{th} column of the product C ($\mathbf{c}_j$) is obtained by multiplying the j^{th} column of the matrix B ($\mathbf{b}_j$) by the matrix A, that is:

$$\mathbf{c}_j = A\mathbf{b}_j, \text{ for all } j = 1, \cdots, n.$$

Thus we can construct the product, C, column by column, just repeating what we have done before, i.e. multiplying a column vector by a matrix .

Examples 11

Let us multiply $\begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{pmatrix}$, which is a (3×2) matrix, and $\begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix}$, which is a (2×2) matrix.

$$\begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{pmatrix}\begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} = \left(\begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{pmatrix}\begin{pmatrix} 3 \\ -4 \end{pmatrix}\right)$$

$$= \begin{pmatrix} (1)(-1)+(2)(2) & (1)(3)+(2)(-4) \\ (3)(-1)+(5)(2) & (3)(3)+(5)(-4) \\ (8)(-1)+(7)(2) & (8)(3)+(7)(-4) \end{pmatrix}$$

$$= \begin{pmatrix} 3 & -5 \\ 7 & -11 \\ 6 & -4 \end{pmatrix}.$$

Let us multiply $\begin{pmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{pmatrix}$, which is a (2×2) matrix, and $\begin{pmatrix} 1+i & -2+i \\ -1+i & 3+2i \end{pmatrix}$,

a is a (2×2) matrix. Note the sizes of these two matrices are compatible.

$$\begin{pmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{pmatrix}\begin{pmatrix} 1+i & -2+i \\ -1+i & 3+2i \end{pmatrix}$$

$$= \left(\begin{pmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{pmatrix}\begin{pmatrix} 1+i \\ -1+i \end{pmatrix}, \begin{pmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{pmatrix}\begin{pmatrix} -2+i \\ 3+2i \end{pmatrix}\right)$$

$$= \begin{pmatrix} (2-i)(1+i)+(1-2i)(-1+i) & (2-i)(-2+i)+(1-2i)(3+2i) \\ (3-2i)(1+i)+(1-3i)(-1+i) & (3-2i)(-2+i)+(1-3i)(3+2i) \end{pmatrix}$$

$$= \begin{pmatrix} 4+4i & 4 \\ 7+5i & 5 \end{pmatrix}.$$

Note that the two matrices must be of compatible sizes for the possibility of matrix multiplication to exist. For example, if $A \in M_{2\times 3}$ and $B \in M_{3\times 5}$ then the product AB is defined, however the product BA, is not defined.

Usually we construct the product $AB = C$ one term at a time.

$(C)_{ij}$ is the entry in the i^{th} row and in the j^{th} column of the product: $(C)_{ij} = (\mathbf{c}_j)_i$.
We know from our earlier work, that the i^{th} entry of $(A\mathbf{b}_j)$ is $\sum\limits_{k=1}^{n} \left(\mathbf{A}^i\right)_k (\mathbf{b}_j)_k$.
Thus we have:

$$(C)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj} = \sum_{k=1}^{n} \left(\mathbf{A}^i\right)_k (\mathbf{b}_j)_k.$$

It is important to remember that in order to obtain the $(i,j)^{th}$ entry of the product of A and B you need to multiply the corresponding entries of the i^{th} row of A and the j^{th} column of B, and no other terms are needed.

Example 12

Suppose that we just want the $(2,2)^{th}$ entry of the product $C = \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{pmatrix}\begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix}$

then we have $c_{22} = \begin{pmatrix} 3 & 5 \end{pmatrix}\begin{pmatrix} 3 \\ -4 \end{pmatrix} = 9 - 20 = -11$.

Suppose that we just want the $(2,1)^{th}$ of the matrix D given by

$$D = \begin{pmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{pmatrix} \begin{pmatrix} 1+i & -2+i \\ -1+i & 3+2i \end{pmatrix}, \text{ then}$$

$$d_{21} = \begin{pmatrix} 3-2i & 1-3i \end{pmatrix} \begin{pmatrix} 1+i \\ -1+i \end{pmatrix} = (3-2i)(1+i) + (1-3i)(-1+i) = 7+5i.$$

We already know that when we evaluate $A\mathbf{x}$, the resulting column vector is a linear combination of the columns of the matrix A. When we evaluate a matrix product of the form AB, then each column of the resulting matrix is a linear combination of the columns of the matrix A. Thus the columns of the matrix A, and the vectors that can be constructed out of these columns using linear combinations, are very important.

Definition 8: Column space.

We define the **column space** of a matrix A, $Col(A)$, to be the span of the columns of A :

$$Col(A) = Span(\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\})$$

$$Col(A) = \{\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \cdots + \alpha_n \mathbf{a}_n : \ \alpha_1, \cdots, \alpha_n \in \mathbb{F}\}.$$

It follows that if $A \in M_{m\times n}$ and $B \in M_{n\times p}$, so that the product, $AB = C$, is defined, then each column of C lies in $Col(A)$:

$$\mathbf{c}_k \in Col(A), \ \text{for } k = 1, \cdots, p.$$

Example 13

Let us extend Example 11.

Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{pmatrix}$, $B = \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{pmatrix} \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} = \begin{pmatrix} 3 & -5 \\ 7 & -11 \\ 6 & -4 \end{pmatrix}$.

Then

$$Col(A) = \{\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 : \ \alpha_1, \alpha_2 \in \mathbb{R}\} = \left\{ \alpha_1 \begin{pmatrix} 1 \\ 3 \\ 8 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 5 \\ 7 \end{pmatrix} : \alpha_1, \alpha_2 \in \mathbb{R} \right\}.$$

And

$$\mathbf{c}_1 = \begin{pmatrix} 3 \\ 7 \\ 6 \end{pmatrix} = (-1) \begin{pmatrix} 1 \\ 3 \\ 8 \end{pmatrix} + (2) \begin{pmatrix} 2 \\ 5 \\ 7 \end{pmatrix} \in Col(A)$$

$$\mathbf{c}_2 = \begin{pmatrix} -5 \\ -11 \\ -4 \end{pmatrix} = (3)\begin{pmatrix} 1 \\ 3 \\ 8 \end{pmatrix} + (-4)\begin{pmatrix} 2 \\ 5 \\ 7 \end{pmatrix} \in Col(A).$$

Lemma 6: Solution of a linear system.

The system of linear equations $A\mathbf{x} = \mathbf{b}$, has a solution *iff* $\mathbf{b} \in Col(A)$.

Proof

a) We first suppose that $A\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$. That is:

$$(\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)(x_1, x_2, \cdots, x_n)^T = \mathbf{b},$$

or equivalently,

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

Thus, since $\mathbf{b}$ is a linear combination of the columns of the matrix A, then $\mathbf{b} \in Col(A)$.

b) We now suppose that $\mathbf{b} \in Col(A)$. We can then write:

$$\mathbf{b} = y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \cdots + y_n\mathbf{a}_n\,, \text{ for some scalars } y_1, y_2, \cdots, y_n.$$

If we let $\mathbf{x} = (y_1, y_2, \cdots, y_n)^T$, then we have:

$$A\mathbf{x} = y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \cdots + y_n\mathbf{a}_n = \mathbf{b},$$

that is $A\mathbf{x} = \mathbf{b}$, and thus we show that this system of linear equations has a solution $\mathbf{x}$.

We conclude then that the system of linear equations is consistent iff $\mathbf{b} \in Col(A)$. $\blacksquare$

Topic 12A

Some Properties of Matrices

Definition 1: Equality.

Let $A \in M_{m\times n}$ and $B \in M_{p\times q}$, then we say that A and B are equal to mean that

(i) $m = p$ and $n = q$, that is, the matrices are the same size, and

(ii) $a_{ij} = b_{ij}$, for $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$, i.e. the corresponding entries are equal.

The set $M_{m\times n}$, of all m by n matrices, is very similar to $\mathbb{R}^n$ and $\mathbb{C}^n$ in that the following properties hold.

Definition 2: Addition.

If both $A, B \in M_{m\times n}$, then we can define addition componentwise by:

(i) $D = A + B\,, \quad D \in M_{m\times n}\,,$ and

(ii) $d_{ij} = a_{ij} + b_{ij}$ for all $i = 1, \cdots, m$ and $j = 1, \cdots, n$.

Addition of matrices that do not have the same size is **not** defined.

Addition satisfies the following, unsurprising rules.

Lemma 1: Properties of matrix addition.

Let A, B and $C \in M_{m\times n}$, then:

$A + B = B + A$

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

There exists a zero matrix, $\mathbb{O} \in M_{m\times n}$ and $\mathbb{O} + A = A + \mathbb{O} = A$.

We define the matrix $-A$, by $(-A)_{ij} = -a_{ij}$, for all $i = 1, \cdots, m$ and $j = 1, \cdots, n$ such that $A + (-A) = (-A) + A = \mathbb{O}$.

It is important that the zero matrix is distinguished from the zero number: if typed, we use the notation $\mathbb{O}$, if writing on paper, it can be written as a number zero with a single vertical bar in the inside left side.

Usually, the context in which the zero matrix is used is sufficient for us to determine the size of the zero matrix. However, when it is not clear which size of the zero matrix is being used, its size is then specified by writing the zero matrix as $\mathbb{O}_{m\times n}$ to avoid confusion.

Example 1

$$\mathbb{O}_{3\times3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbb{O}_{2\times3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbb{O}_{3\times2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Definition 3: Multiplication of a matrix by a scalar.

Let $A \in M_{m\times n}$ and $c \in \mathbb{F}$, then:

$F = cA$ and $F \in M_{m\times n}$

such that $f_{ij} = ca_{ij}$, for all $i = 1, \cdots, m$ and $j = 1, \cdots, n$.

Lemma 2: Properties of multiplication of a matrix by a scalar.

Let $A, B \in M_{m\times n}$, $C \in M_{n\times p}$ and $c, d \in \mathbb{F}$, we then have:

$cA = Ac$

$c(A + B) = cA + cB$

$(c + d)A = cA + dA$

$c(dA) = (cd)A$

$c(AC) = (cA)C = A(cC) = cAC$

Definition 4: Transpose of a matrix.

Let $A \in M_{m\times n}$. We define the transpose of A, denoted A^T, by $\left(A^T\right)_{pq} = (A)_{qp}$.

Notice that $A^T \in M_{n\times m}$ and so this is a matrix with n rows and m columns. It is formed by making all the rows of A into the columns of A^T in the order in which they appear.

Example 2

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \end{pmatrix}, \text{ then } A^T = \begin{pmatrix} 1 & 4 \\ 2 & -5 \\ -3 & 6 \end{pmatrix}.$$

Lemma 3: Properties of the transpose.

If $A, B \in M_{m\times n}$ and if $c \in \mathbb{F}$, then:

$(A+B)^T = A^T + B^T$

$(cA)^T = cA^T$

$\left(A^T\right)^T = A$

Lemma 4: Properties of matrix multiplication.

Let $A, G \in M_{m\times n}$, $B, D \in M_{n\times p}$ and $C \in M_{p\times q}$, then:

$(A+G)B = AB + GB$

$A(B+D) = AB + AD$

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

$(AB)^T = B^T A^T$, the transpose of a product is the product of the transposes IN REVERSE.

Warning: Let $A \in M_{m\times n}$ and $B \in M_{n\times q}$. If $m \neq q$, then "BA" is **not defined**.
If $m = q$, so that BA is defined, there is no reason why it should equal AB,
in fact the two matrices might not even be the same size.

Example 3

Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then:

$AB = \begin{pmatrix} 4 & 6 \\ 0 & 0 \end{pmatrix}$ and $BA = \begin{pmatrix} 1 & 1 \\ 3 & 3 \end{pmatrix}$.

Clearly, although both products AB and BA are defined, $AB \neq BA$.

Note: if a, b, $c \in \mathbb{F}$, then we have two important results, namely:

(i) If $ab = ac$ and $a \neq 0$, then $b = c$ (cancellation law)
(ii) $ab = 0$ iff either $a = 0$ or $b = 0$

These results do not extend to matrix arithmetic in general.

Example 4

Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix}$, and $C = \begin{pmatrix} 2 & 5 \\ 3 & 4 \end{pmatrix}$. Then,

$AB = AC = \begin{pmatrix} 3 & 4 \\ 6 & 8 \end{pmatrix}$ and $A \neq \mathbb{O}$, but $B \neq C$.

Example 5

Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 7 \\ 0 & 0 \end{pmatrix}$, Then,

$AB = \mathbb{O}$, but $A \neq \mathbb{O}$ and $B \neq \mathbb{O}$.

Topic 12B

Special Square Matrices

Definition 5: Square matrix.

Let $A \in M_{m\times n}$, we say that A is a square matrix to mean that $m = n$.

Example 6

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -2+i & 3-2i \\ 2+2i & 6-7i \end{pmatrix}.$$

The following definitions only apply to square matrices.

Definition 6: Symmetric and skew-symmetric.

Let $B \in M_{n\times n}$, we say that B is **symmetric** to mean that $B = B^T$.

Let $C \in M_{n\times n}$ we say that C is **skew-symmetric** to mean that $C = -C^T$.

Example 7

$$\begin{pmatrix} 1 & -7 \\ -7 & 5 \end{pmatrix} \text{ and } \begin{pmatrix} 2 & 4 & 6 \\ 4 & -5 & 7 \\ 6 & 7 & 9 \end{pmatrix} \text{ are symmetric.}$$

$$\begin{pmatrix} 0 & -7 \\ 7 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & -4 & 6 \\ 4 & 0 & 7 \\ -6 & -7 & 0 \end{pmatrix} \text{ are skew-symmetric.}$$

Definition 7: Upper Triangular

Let $A \in M_{n\times n}$, we say that A is upper triangular $(U\Delta)$ to mean that $a_{ij} = 0$ for $i > j$ with $i,\, j = 1, \cdots, n$.

Example 8

$$\begin{pmatrix} 1+i & -7 \\ 0 & 1-i \end{pmatrix}, \begin{pmatrix} 3 & -4 & 6 \\ 0 & 5 & 9 \\ 0 & 0 & 7 \end{pmatrix}, \text{ and } \begin{pmatrix} i & 4-i & 6-4i \\ 0 & 1+i & 7i \\ 0 & 0 & 2-4i \end{pmatrix} \text{ are upper triangular.}$$

Definition 8: Lower Triangular

Let $B \in M_{n\times n}$, we say that B is lower triangular $(L\Delta)$ to mean that $b_{ij} = 0$ for $i < j$ with $i,\, j = 1, \cdots, n$.

Example 9

$$\begin{pmatrix} i & 0 \\ 4 & 1-i \end{pmatrix}, \begin{pmatrix} 3 & 0 & 0 \\ 3 & 4 & 0 \\ 0 & 7 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} i & 0 & 0 & 0 \\ 7i & 1+i & 0 & 0 \\ 6 & 2+i & 2-4i & 0 \\ 3-5i & 5 & 0 & 4+3i \end{pmatrix} \text{ are lower triangular.}$$

Remarks:

The transpose of an upper (lower) triangular matrix is a lower (upper) triangular matrix, respectively.
The product of upper (lower) triangular matrices is upper (lower) triangular, respectively.

Definition 9: Diagonal

Let $C \in M_{n\times n}$, we say that C is diagonal to mean that $c_{ij} = 0$ for $i \neq j$ with $i,\, j = 1, \cdots, n$.

Example 10

$$\begin{pmatrix} i & 0 \\ 0 & 1-i \end{pmatrix}, \begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \text{ and } \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 2-4i & 0 \\ 0 & 0 & 0 & 4+3i \end{pmatrix} \text{ are diagonal.}$$

Notice that a diagonal matrix is both upper triangular and lower triangular.
The zero matrix is diagonal, and thus upper and lower triangular.

Definition 10: Diagonal entries and main diagonal

Let $A \in M_{n\times n}$. We refer to the entries a_{ii} as the diagonal entries of A, and we call $(a_{11}, a_{22}, \cdots, a_{nn})$, the main diagonal of A.

When $C \in M_{n\times n}$ is a diagonal matrix, then it is only the entries on the main diagonal which are possibly non-zero. In this case, we can specify the matrix C by just providing its diagonal entries, and we write $C = diag(c_{11}, c_{22}, \cdots, c_{nn})$.

Example 11

$$C = diag(1,-2,3) \text{ means that } C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

We will see that diagonal matrices are very important matrices in Linear Algebra.

Definition 11: Identity matrix

The diagonal matrix, $diag(1,1,\cdots,1)$ is called the identity matrix, and is denoted by I. If we wish to indicate the size of the identity matrix we add a subscript n.

The diagonal matrix I_n plays a role for matrices that is similar to the role played by the number 1 for scalars, in that it is a multiplicative identity, that is, if $A \in M_{m\times n}$ then:

$$I_m A = A \text{ and } AI_n = A.$$

Example 12

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

If $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 0 \end{pmatrix}$, then

$$AI_2 = I_2 A = A\ ,\ BI_3 = I_2 B = B\ , \quad \text{and} \quad cI_n = diag(c,\, c, \cdots, c) \ \text{ for } c \in \mathbb{F}.$$

Definition 12: Elementary matrices

We say that a matrix is an **elementary matrix** to mean that it is obtained from the identity matrix by performing a **single** elementary row operation.

If we wish to be more precise, we refer to elementary matrices of type I, II and III, according to which type of elementary row operation is performed.

Example 13

$$E_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ is a type I elementary matrix } (R_1 \leftrightarrow R_3),$$

$$\text{since } \ I_3 \xrightarrow{R_1 \leftrightarrow R_3} \ = E_1.$$

$$E_2 = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} \text{ is a type II elementary matrix } (R_1 \to 5R_1),$$

$$\text{since } I_2 \xrightarrow{R_1 \to 5R_1} = E_2.$$

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \text{ is a type III elementary matrix } (R_2 \to R_2 - 2R_3),$$

$$\text{since } I_3 \xrightarrow{R_2 \to R_2 - 2R_3} = E_3.$$

Elementary matrices are useful because they store the information of the steps involved in performing elementary row operations, and in addition we have the following result:

Lemma 5

Let $C \in M_{m \times n}$ and suppose that a single elementary row operation is performed on it to produce a matrix B. Suppose, also, that we perform the same elementary row operation on the matrix I_m to produce the elementary matrix E. Then:

$$B = EC.$$

This means that we can either just perform the elementary row operation on C to get the matrix B, or we can store the information of that row operation in the elementary matrix E, instead.

Lemma 6: Induction on Lemma 5.

Let $C \in M_{m \times n}$ and suppose that a finite number, q, of elementary row operations, labelled $op(1)$, $op(2), \cdots, op(q)$, are performed on C to produce a matrix D.

Suppose, also, that when we perform $op(1)$ on I_m we produce the elementary matrix E_1, and when we perform $op(2)$ on I_m we produce the elementary matrix $E_2, \ldots,$
and when we perform $op(q)$ on I_m we produce the elementary matrix E_q, then:

$$D = E_q \ldots E_2 E_1 C$$

Notice the order in which the elementary matrices appear in this product.

Example 14

Let us revisit Example 1 in Topic 10A. The augmented matrix, B, was given by

$$B = \left(\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 2 & -4 & 1 & 0 & 5 \\ 1 & -2 & 2 & -3 & 4 \end{array} \right).$$

We then performed the following elementary row operations:

$$\xrightarrow[R3\to R3-R1]{R2\to R2-2R1} \begin{pmatrix} 1 & -2 & -1 & 3 & | \, 1 \\ 0 & 0 & 3 & -6 & | \, 3 \\ 0 & 0 & 3 & -6 & | \, 3 \end{pmatrix} \quad (= C)$$

$$\xrightarrow{R_2\to \frac{1}{3}R2} \begin{pmatrix} 1 & -2 & -1 & 3 & | \, 1 \\ 0 & 0 & 1 & -2 & | \, 1 \\ 0 & 0 & 3 & -6 & | \, 3 \end{pmatrix} \quad (= D)$$

$$\xrightarrow{R3\to R3-3R2} \begin{pmatrix} 1 & -2 & -1 & 3 & | \, 1 \\ 0 & 0 & 1 & -2 & | \, 1 \\ 0 & 0 & 0 & 0 & | \, 0 \end{pmatrix} = F.$$

We write down the four corresponding elementary matrices:

$$I_3 \xrightarrow{R_2\to R_2-2R_1} = E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad I_3 \xrightarrow{R_3\to R_3-R_1} = E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix},$$

(and thus $C = E_2E_1B$)

$$I_3 \xrightarrow{R_2\to \frac{1}{3}R_2} = E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad I_3 \xrightarrow{R_3\to R_3-3R_2} = E_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix},$$

(thus $D = E_3C$ and $F = E_4D$).

We then evaluate the product: $E_4E_3E_2E_1B = E_4(E_3(E_2(E_1B)))$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -1 & 3 & | \, 1 \\ 2 & -4 & 1 & 0 & | \, 5 \\ 1 & -2 & 2 & -3 & | \, 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -1 & 3 & | \, 1 \\ 0 & 0 & 3 & -6 & | \, 3 \\ 1 & -2 & 2 & -3 & | \, 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -1 & 3 & | \, 1 \\ 0 & 0 & 3 & -6 & | \, 3 \\ 0 & 0 & 3 & -6 & | \, 3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & -1 & 3 & | \, 1 \\ 0 & 0 & 1 & -2 & | \, 1 \\ 0 & 0 & 3 & -6 & | \, 3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -2 & -1 & 3 & |1 \\ 0 & 0 & 1 & -2 & |1 \\ 0 & 0 & 0 & 0 & |0 \end{pmatrix} = F$$

Topic 13A

Linear Transformations I

Up to this point in the course, we have considered matrices as arrays of numbers which contain important information about systems of linear equations through either the coefficient matrix, the augmented matrix, or both.
Moving forward we will discover that there is much more going on than this.
We have defined matrix multiplication in the previous lectures, and we will make use of that operation to define functions.

Definition 1: Function determined by a matrix.

Let $A \in M_{m\times n}(\mathbb{F})$. We then define **the function determined by the matrix** A, denoted T_A by

$$T_A : \mathbb{F}^n \to \mathbb{F}^m, \quad \text{with } \; T_A(\mathbf{x}) = A\mathbf{x}.$$

Examples 1

Let $B \in M_{3\times 2}(\mathbb{R})$, and $B = \begin{pmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{pmatrix}$. If $\mathbf{x} \in \mathbb{R}^2,$ let us define $T_B(\mathbf{x})$ by

$$T_B(\mathbf{x}) = \begin{pmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 4x_2 \\ -2x_1 - 5x_2 \\ 4x_1 + 6x_2 \end{pmatrix}.$$

Note that $T_B(\mathbf{x}) \in \mathbb{R}^3$.

For instance, if $\mathbf{x} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$, then $T_B \begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} 10 \\ -11 \\ 10 \end{pmatrix}$.

The function T_B, determined by the matrix B, converts vectors in $\mathbb{R}^2$ into vectors in $\mathbb{R}^3$.

Let $G \in M_{2\times 3}(\mathbb{C})$, and $G = \begin{pmatrix} i & 1+2i & 3+2i \\ 2-i & 4 & 2-5i \end{pmatrix}$. If $\mathbf{z} \in \mathbb{C}^3,$ let us define $T_G(\mathbf{z})$ by

$$T_G(\mathbf{z}) = \begin{pmatrix} i & 1+2i & 3+2i \\ 2-i & 4 & 2-5i \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} iz_1 + (1+2i)z_2 + (3+2i)z_3 \\ (2-i)z_1 + 4z_2 + (2-5i)z_3 \end{pmatrix}.$$

Note that $T_G(\mathbf{z}) \in \mathbb{C}^2$.

For instance, if $\mathbf{z} = \begin{pmatrix} 3 \\ 2-i \\ 2i \end{pmatrix}$, then $T_G\left(\begin{pmatrix} 3 \\ 2-i \\ 2i \end{pmatrix}\right) = \begin{pmatrix} 12i \\ 24-3i \end{pmatrix} = 3\begin{pmatrix} 4i \\ 8-i \end{pmatrix}$.

The function T_G, determined by the matrix G converts vectors in $\mathbb{C}^3$ into vectors in $\mathbb{C}^2$.

The function determined by a matrix has some rather special features.
In particular, we recall Lemma 1 (T11A), concerning the linearity of matrix multiplication: that is,

Let $A \in M_{m\times n}(\mathbb{F})$, $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, and $c \in \mathbb{F}$, then: $\begin{cases} A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} \\ A(c\mathbf{x}) = cA\mathbf{x} \end{cases}$.

Lemma 1

Let $A \in M_{m\times n}(\mathbb{F})$, $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, and $c \in \mathbb{F}$.
Let T_A be the function determined by the matrix A, then T_A is linear, that is:

$$\begin{cases} T_A(\mathbf{x}+\mathbf{y}) = T_A(\mathbf{x}) + T_A(\mathbf{y}) \\ T_A(c\mathbf{x}) = cT_A(\mathbf{x}) \end{cases}$$

Proof

The proof of this result follows at once from Lemma 1 (T11A), and Definition 1:

$$\begin{aligned} T_A(\mathbf{x}+\mathbf{y}) &= A(\mathbf{x}+\mathbf{y}) & \quad \textbf{and} \quad && T_A(c\mathbf{x}) &= A(c\mathbf{x}) && \text{by definition} \\ &= A(\mathbf{x}) + A(\mathbf{y}) & && &= cA\mathbf{x} && \text{by Lemma 1 (T11A)} \\ &= T_A(\mathbf{x}) + T_A(\mathbf{y}) & && &= cT_A(\mathbf{x}) && \text{by definition.} \quad \blacksquare \end{aligned}$$

Let us now consider functions from $\mathbb{F}^n$ to $\mathbb{F}^m$, $T : \mathbb{F}^n \to \mathbb{F}^m$. In linear algebra, we only consider a very special type of function known as a linear function or linear transformation.

Definition 2: Linear Transformation.

Let $T : \mathbb{F}^n \to \mathbb{F}^m$, we say that T is a **linear transformation** to mean that, for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ and $c \in \mathbb{F}$,

$$\begin{cases} T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) \ , \quad \text{this is called linearity over addition} \\ T(c\mathbf{x}) = cT(\mathbf{x}) \ , \quad \text{this is called linearity over scalar multiplication.} \end{cases}$$

Note that in linear algebra, the word **transformation** is often used instead of function.

Clearly, if $A \in M_{m\times n}(\mathbb{F})$, and T_A is the function determined by A, then T_A is a linear transformation. That is, the functions which are determined by matrices are rather special.

Some of you may be a little unsettled by the fact that we seem to be entering a strange new world where we examine functions $T : \mathbb{F}^n \to \mathbb{F}^m$. Let me try to allay your fears.

In first year calculus, you deal with functions, $f : A \to B$, where the domain and codomain are very simple, namely subsets of $\mathbb{R}$. However the functions can be **quite complicated.**
In linear algebra, our domains and codomains will be different than those in first year calculus, namely $\mathbb{R}^n$ and $\mathbb{C}^n$. However, whenever we are considering functions, we will restrict our attention almost exclusively to linear transformations. Thus we have more interesting spaces but much simpler functions.
FYI, the two arenas of **more complicated spaces** and **more complicated functions** are seen together in a vector calculus course.

We will now spend some time considering some features of linear transformations.

Lemma 2: Alternate characterization of a linear transformation.

Let $T : \mathbb{F}^n \to \mathbb{F}^m$.

T is a **linear transformation** iff $T(c_1\mathbf{x} + c_2\mathbf{y})=c_1T(\mathbf{x}) + c_2T(\mathbf{y})$,

for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ and for all $c_1, c_2 \in \mathbb{F}$.

Proof

If $c_1, c_2 \in \mathbb{F}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, then $c_1\mathbf{x} = \mathbf{z}_1 \in \mathbb{F}^n$ and $c_2\mathbf{y} = \mathbf{z}_2 \in \mathbb{F}^n$.

If T is a linear transformation, then: $T(\mathbf{z}_1 + \mathbf{z}_2) = T(\mathbf{z}_1) + T(\mathbf{z}_2)$.
Also, since T is a linear transformation, then: $T(\mathbf{z}_1) = T(c_1\mathbf{x}) = c_1T(\mathbf{x})$ and $T(\mathbf{z}_2) = c_2T(\mathbf{y})$.
Thus we have: $T(c_1\mathbf{x} + c_2\mathbf{y}) = T(\mathbf{z}_1 + \mathbf{z}_2) = T(\mathbf{z}_1) + T(\mathbf{z}_2) = c_1T(\mathbf{x}) + c_2T(\mathbf{y})$.

On the other hand, we now suppose that $T(c_1\mathbf{x} + c_2\mathbf{y})=c_1T(\mathbf{x}) + c_2T(\mathbf{y})$,
for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, and for all $c_1, c_2 \in \mathbb{F}$.

In particular, if $c_1 = c_2 = 1$, we have: $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$,
and also if $c_2 = 0$, we have: $T(c_1\mathbf{x}) = c_1T(\mathbf{x})$ for all $\mathbf{x}$, and for all $c_1 \in \mathbb{F}$.

We conclude that T is a linear transformation. $\blacksquare$

This result means that, if we want, we can check so see if a function is linear, "all in one go," instead of checking linearity over addition and then linearity over scalar multiplication separately. We will have a few examples, after the next result.

Lemma 3: The zero vector is mapped to the zero vector.

Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation, then $T(\mathbf{0}_{\mathbb{F}^n}) = \mathbf{0}_{\mathbb{F}^m}$.

Proof:

Let $T(\mathbf{0}_{\mathbb{F}^n}) = \mathbf{y} \in \mathbb{F}^m$. We will show that $\mathbf{y} = \mathbf{0}_{\mathbb{F}^m}$.

Recall that $0(\mathbf{0}_{\mathbb{F}^n}) = \mathbf{0}_{\mathbb{F}^n}$. Since T is a linear transformation, we have that:

$\mathbf{y} = T(\mathbf{0}_{\mathbb{F}^n}) = T(0(\mathbf{0}_{\mathbb{F}^n})) = 0T(\mathbf{0}_{\mathbb{F}^n}) = 0\mathbf{y} = \mathbf{0}_{\mathbb{F}^m}$, using properties of the scalar zero.
That is, $\mathbf{y} = \mathbf{0}_{\mathbb{F}^m}$.

Thus, a linear transformation maps the zero vector of $\mathbb{F}^n$ to the zero vector of $\mathbb{F}^m$. ■

We will now look at some functions from $\mathbb{F}^n \to \mathbb{F}^m$, and determine if they are linear or not. The key idea is to first make a decision about what you think is going to happen.

Does the zero vector of the domain get mapped to the zero vector of the codomain?
If it does not, then the function is not linear.

Does the function look linear? If it does not, then a simple counterexample is needed to prove that the function is not linear.

If the answer to both of these questions is yes, then use the definition to prove that it is linear.

Example 2

Let $T_1 : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T_1\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} 2x + 3y + 4 \\ 6x - 7z \end{pmatrix}$.

This transformation is not linear because $T_1((0, 0, 0)^T) = (4, 0)^T \neq (0, 0)^T$.

Example 3

Let $T_2 : \mathbb{C}^3 \to \mathbb{C}^2$ be defined by $T_2\left(\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}\right) = \begin{pmatrix} 2z_1 + 3z_3 \\ z_2 z_3 \end{pmatrix}$.

The zero vector of $\mathbb{C}^3$ is mapped to the zero vector of $\mathbb{C}^2$ by this function. However, this function does not look linear because of the product term $z_2 z_3$.

We try to construct a counterexample. The (possible) troublesome term does not involve z_1, so we will set that to zero. Since the image under the function does involve the other two variables let us see what happens for different values (setting both to 1 or 0 is not usually helpful).

$$T_2\left(\begin{pmatrix}0\\2\\2\end{pmatrix}\right)=\begin{pmatrix}6\\4\end{pmatrix},\ T_2\left(\begin{pmatrix}0\\3\\3\end{pmatrix}\right)=\begin{pmatrix}9\\9\end{pmatrix},\text{ and, }\begin{pmatrix}0\\2\\2\end{pmatrix}+\begin{pmatrix}0\\3\\3\end{pmatrix}=\begin{pmatrix}0\\5\\5\end{pmatrix}.$$

However, $T_2\left(\begin{pmatrix}0\\5\\5\end{pmatrix}\right)=\begin{pmatrix}15\\25\end{pmatrix}$, and yet $\begin{pmatrix}6\\4\end{pmatrix}+\begin{pmatrix}9\\9\end{pmatrix}=\begin{pmatrix}15\\13\end{pmatrix}$.

We then have: $T_2\left(\begin{pmatrix}0\\2\\2\end{pmatrix}+\begin{pmatrix}0\\3\\3\end{pmatrix}\right)\neq T_2\left(\begin{pmatrix}0\\2\\2\end{pmatrix}\right)+T_2\left(\begin{pmatrix}0\\3\\3\end{pmatrix}\right)$.

The transformation T_2 is not a linear transformation.

We will not be dealing with functions like T_2 in this course.

Example 4

$T_3:\mathbb{R}^2\to\mathbb{R}^3$ be defined by $T_3\left(\begin{pmatrix}x\\y\end{pmatrix}\right)=\begin{pmatrix}2x+3y\\6x-5y\\2x\end{pmatrix}$.

This could be linear, the zero vector of $\mathbb{R}^2$ is mapped to the zero vector of $\mathbb{R}^3$, and we do not detect any "troublesome" terms. We will prove that T_3 is linear.

Let $\mathbf{x}=\begin{pmatrix}x_1\\x_2\end{pmatrix}$, $\mathbf{y}=\begin{pmatrix}y_1\\y_2\end{pmatrix}\in\mathbb{R}^2$, and $a,b\in\mathbb{R}$. We then have, using the definition of T_3:

$$T_3(\mathbf{x})=T_3\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\right)=\begin{pmatrix}2x_1+3x_2\\6x_1-5x_2\\2x_1\end{pmatrix},\text{ and }T_3(\mathbf{y})=T_3\left(\begin{pmatrix}y_1\\y_2\end{pmatrix}\right)=\begin{pmatrix}2y_1+3y_2\\6y_1-5y_2\\2y_1\end{pmatrix}.$$

We also have : $a\mathbf{x}+b\mathbf{y}=\begin{pmatrix}ax_1\\ax_2\end{pmatrix}+\begin{pmatrix}by_1\\by_2\end{pmatrix}=\begin{pmatrix}ax_1+by_1\\ax_2+by_2\end{pmatrix}$,

$$\text{so that } T_3\left(a\mathbf{x}+b\mathbf{y}\right)=T_3\left(\begin{pmatrix}ax_1+by_1\\ax_2+by_2\end{pmatrix}\right)=\begin{pmatrix}2(ax_1+by_1)+3(ax_2+by_2)\\6(ax_1+by_1)-5(ax_2+by_2)\\2(ax_1+by_1)\end{pmatrix}$$

$$=a\begin{pmatrix}2x_1+3x_2\\6x_1-5x_2\\2x_1\end{pmatrix}+b\begin{pmatrix}2y_1+3y_2\\6y_1-5y_2\\2y_1\end{pmatrix}=aT_3\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}\right)+bT_3\left(\begin{pmatrix}y_1\\y_2\end{pmatrix}\right)$$

that is, $T_3\left(a\mathbf{x}+b\mathbf{y}\right)=aT_3\left(\mathbf{x}\right)+bT_3\left(\mathbf{y}\right)$. (This is true for all $\mathbf{x}$, $\mathbf{y}\in\mathbb{R}^2$ and for all a, $b\in\mathbb{R}$).

We thus conclude that T_3 is a linear transformation.

Example 5

$T_4 : \mathbb{C}^2 \to \mathbb{C}^2$ be defined by $T_4\left(\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) = \begin{pmatrix} 2iz_1 + 3z_2 \\ (2-i)z_1 - (1+3i)z_2 \end{pmatrix}$.

This could be linear, the zero vector of $\mathbb{C}^2$ is mapped to the zero vector of $\mathbb{C}^2$, and we do not detect any "troublesome" terms. We will prove that it is linear.

Let $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \in \mathbb{C}^2$, and $a, b \in \mathbb{C}$.

We then have, from the definition of T_4:

$$T_4(\mathbf{z}) = T_4\left(\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) = \begin{pmatrix} 2iz_1 + 3z_2 \\ (2-i)z_1 - (1+3i)z_2 \end{pmatrix}, \text{ and}$$

$$T_4(\mathbf{w}) = T_4\left(\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}\right) = \begin{pmatrix} 2iw_1 + 3w_2 \\ (2-i)w_1 - (1+3i)w_2 \end{pmatrix}.$$

We also have: $a\mathbf{z} + b\mathbf{w} = \begin{pmatrix} az_1 \\ az_2 \end{pmatrix} + \begin{pmatrix} bw_1 \\ bw_2 \end{pmatrix} = \begin{pmatrix} az_1 + bw_1 \\ az_2 + bw_2 \end{pmatrix}$,

so that
$$\begin{aligned} T_4((a\mathbf{z} + b\mathbf{w})) = T_4\left(\begin{pmatrix} az_1 + bw_1 \\ az_2 + bw_2 \end{pmatrix}\right) &= \begin{pmatrix} 2i(az_1 + bw_1) + 3(az_2 + bw_2) \\ (2-i)(az_1 + bw_1) - (1+3i)(az_2 + bw_2) \end{pmatrix} \\ &= a\begin{pmatrix} 2iz_1 + 3z_2 \\ (2-i)z_1 - (1+3i)z_2 \end{pmatrix} + b\begin{pmatrix} 2iw_1 + 3w_2 \\ (2-i)w_1 - (1+3i)w_2 \end{pmatrix} \\ &= aT_4\left(\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) + bT_4\left(\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}\right), \end{aligned}$$

that is, $T_4((a\mathbf{z} + b\mathbf{w})) = aT_4(\mathbf{z}) + bT_4(\mathbf{w})$. (This is true for all $\mathbf{z}$, $\mathbf{w} \in \mathbb{C}^2$ and for all a, $b \in \mathbb{F}$).

We conclude that T_4 is a linear transformation.

A linear transformation is a special case of a function, and, as with all functions, we are interested in the set of image points and the set of zeros of the function.

Definition 3: Range.

Let $T : \mathbb{F}^n \to \mathbb{F}^m$, we define the **range** of T, denoted $R(T)$, to mean the set of image points of T:

$$R(T) = \{T(\mathbf{x}) : \mathbf{x} \in \mathbb{F}^n\}.$$

The range of T is a subset of $\mathbb{F}^m$.

Lemma 4

Let $A \in M_{m\times n}(\mathbb{F})$, and let T_A be the function from $\mathbb{F}^n$ to $\mathbb{F}^m$ determined by the matrix A. Then

$$R(T_A) = Col(A).$$

Proof

We will show that these two subsets of $\mathbb{F}^m$ are equal, by showing that each set is contained in the other.

We first suppose that $\mathbf{y} \in R(T_A)$, then there exists $\mathbf{x} \in \mathbb{F}^n$ such that $T_A(\mathbf{x}) = \mathbf{y}$.
That is, there exists a vector $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{y}$.
We can conclude from Lemma 6 (T11C) that $\mathbf{y} \in Col(A)$.

Now suppose that $\mathbf{z} \in Col(A)$.
There exist then scalars $x_1, x_2, \ldots, x_n \in \mathbb{F}$ such that $\mathbf{z} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$.
If we let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, then $A\mathbf{x} = \mathbf{z}$, that is $\mathbf{z} = T_A(\mathbf{x})$, and so $\mathbf{z} \in R(T_A)$.

We thus conclude that $R(T_A) = Col(A)$. ■

Remark 1: Connection to systems of equations.

We have already seen in Lemma 6 (T11C) that:

the system of linear equations $A\mathbf{x} = \mathbf{b}$ has a solution **iff** $\mathbf{b} \in Col(A)$.

We can now write:

$$A\mathbf{x} = \mathbf{b} \text{ is consistent } \textbf{iff } \mathbf{b} \in R(T_A).$$

We can interpret this in the following way: you can only solve the system of equations $A\mathbf{x} = \mathbf{b}$, when the vector $\mathbf{b}$ lies in the range of the function T_A, determined by A.

Example 6

Let us revisit Example 1. Let $B = \begin{pmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{pmatrix}$ and if $\mathbf{x} \in \mathbb{R}^2$, let us define $T_B(\mathbf{x})$ by

$$T_B(\mathbf{x}) = \begin{pmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 4x_2 \\ -2x_1 - 5x_2 \\ 4x_1 + 6x_2 \end{pmatrix}.$$

Find the range of T_B.

Solution

The range of T_B is easily found to be:

$$R(T_B) = Span\left(\left\{\begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ -5 \\ 6 \end{pmatrix}\right\}\right).$$

Note that you will be able to solve the linear system of equations, $B\mathbf{x} = \mathbf{b}$ iff $\mathbf{b} \in R(T_B)$.

Let $G = \begin{pmatrix} i & 1+2i & 3+2i \\ 2-i & 4 & 2-5i \end{pmatrix}$ and if $\mathbf{z} \in \mathbb{C}^3$, let us define $T_G(\mathbf{z})$ by

$$T_G(\mathbf{z}) = \begin{pmatrix} i & 1+2i & 3+2i \\ 2-i & 4 & 2-5i \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} iz_1 + (1+2i)z_2 + (3+2i)z_3 \\ (2-i)z_1 + 4z_2 + (2-5i)z_3 \end{pmatrix}.$$

Find the range of T_G.

Solution

The range of T_G is easily found to be:

$$R(T_G) = Span\left(\left\{\begin{pmatrix} i \\ 2-i \end{pmatrix}, \begin{pmatrix} 1+2i \\ 4 \end{pmatrix}, \begin{pmatrix} 3+2i \\ 2-5i \end{pmatrix}\right\}\right)$$

It may be shown (try it) that $R(T_G) = \mathbb{C}^2$, and thus it is possible to solve the linear system of equations, $G\mathbf{x} = \mathbf{b}$, for all $\mathbf{b} \in \mathbb{C}^2$.

Definition 4: Onto

We say that the function, $T : \mathbb{F}^n \to \mathbb{F}^m$, is **onto** to mean that $R(T) = \mathbb{F}^m$.

Thus, a function, T, is onto means that the range of T is the entire codomain of T.

In particular, if $S : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation, then we say that S is onto to mean that $R(S) = \mathbb{F}^m$.

Corollary 1: (from Lemma 4)

Let $A \in M_{m\times n}(\mathbb{F})$, and let T_A be the function from $\mathbb{F}^n$ to $\mathbb{F}^m$ determined by the matrix A, then

$$T_A \text{ is onto } \textbf{iff } Col(A) = \mathbb{F}^m.$$

Proof

This follows immediately from Lemma 4, since $R(T_A) = Col(A)$.

Corollary 2: (from Lemma 4)

Let $A \in M_{m\times n}(\mathbb{F})$, and let T_A be the function from $\mathbb{F}^n$ to $\mathbb{F}^m$ determined by the matrix A, then

$$T_A \text{ is onto } \textbf{iff } rank(A) = m.$$

Proof

We know from Corollary 1 that T_A is onto **iff** $Col(A) = \mathbb{F}^m$.

If $Col(A) = \mathbb{F}^m$, then $\mathbf{b} \in \mathbb{F}^m \implies \mathbf{b} \in Col(A)$.
We can then conclude that $A\mathbf{x} = \mathbf{b}$ is consistent and thus

$$rank(A) = rank(A|\mathbf{b}), \forall\, \mathbf{b} \in \mathbb{F}^m.$$

This happens only when $rank(A) = m$, so that the matrix A has a pivot in each row and thus the last column of the augmented matrix $(A|\mathbf{b})$ cannot be a pivot column.

If $rank(A) = m$, then $rank(A) = m = rank(A|\mathbf{b})$ for every $\mathbf{b} \in \mathbb{F}^m$, and thus $A\mathbf{x} = \mathbf{b}$ can be solved for all $\mathbf{b} \in \mathbb{F}^m$, that is, every vector $\mathbf{b} \in \mathbb{F}^m$ satisfies $\mathbf{b} \in Col(A)$.

We then conclude that $Col(A) = \mathbb{F}^m$. ■

Example 7

Let us revisit Example 6.

Let $B = \begin{pmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{pmatrix}$ and if $\mathbf{x} \in \mathbb{R}^2$, let us define $T_B(\mathbf{x})$ by

$$T_B(\mathbf{x}) = \begin{pmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 4x_2 \\ -2x_1 - 5x_2 \\ 4x_1 + 6x_2 \end{pmatrix}.$$

We know from Example 6 that $R(T_B) = Col(B) = Span\left(\left\{\begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 \\ -5 \\ 6 \end{pmatrix}\right\}\right)$, and

thus conclude that T_B is not onto, as $R(T_B) \neq \mathbb{R}^3$, since the range of T_B is only a plane in $\mathbb{R}^3$ and not the whole of $\mathbb{R}^3$.
There will be some vectors, $\mathbf{b}$ in $\mathbb{R}^3$, for which the equation $T_B(\mathbf{x}) = \mathbf{b}$ has no solution.
This will happen for any vector which does not lie on that plane, $Col(B)$.
We will be revisiting this idea later on, once we have talked about dimension.

Definition 5: Nullspace

Let $T : \mathbb{F}^n \to \mathbb{F}^m$, we define the **nullspace** of T, denoted $N(T)$, to mean the set of zeros of T, that is, the set of vectors in $\mathbb{F}^n$ such that their image under T is the zero vector of $\mathbb{F}^m$: i.e.

$$N(T) = \{\mathbf{x} \in \mathbb{F}^n : T(\mathbf{x}) = \mathbf{0}_{\mathbb{F}^m}\}.$$

The nullspace of T is the subset of vectors in $\mathbb{F}^n$, whose image is the zero vector in $\mathbb{F}^m$.

Note that if T is a linear transformation then, by Lemma 3, $\mathbf{0}_{\mathbb{F}^n} \in N(T)$, and so the nullspace of a linear transformation is **never** empty.

Lemma 5

Let $A \in M_{m\times n}(\mathbb{F})$, and let T_A be the function from $\mathbb{F}^n$ to $\mathbb{F}^m$ determined by the matrix A. Then

$$N(T_A) = N(A).$$

Proof

$\mathbf{x} \in N(T_A)$ iff $T_A(\mathbf{x}) = \mathbf{0}_{\mathbb{F}^\mathbf{m}}$ iff $A\mathbf{x} = \mathbf{0}_{\mathbb{F}^m}$ iff $\mathbf{x} \in N(A)$. ■

Definition 6: One-to-one.

We say that the function $T : \mathbb{F}^n \to \mathbb{F}^m$, is **one-to-one** to mean that for every $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$,

$$\text{if } \mathbf{x} \neq \mathbf{y}, \text{ then } T(\mathbf{x}) \neq T(\mathbf{y}).$$

Thus, a function is one-to-one, when distinct points have distinct images.

In particular, if $S : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation, then we say that S is one-to-one to mean that, for every $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$,

$$\text{if } \mathbf{x} \neq \mathbf{y}, \text{ then } S(\mathbf{x}) \neq S(\mathbf{y}).$$

Lemma 6

Let $A \in M_{m\times n}(\mathbb{F})$, and let T_A be the function from $\mathbb{F}^n$ to $\mathbb{F}^m$ determined by the matrix A, then

$$T_A \text{ is one-to-one } \textbf{iff } N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\}.$$

Proof

a) Let T_A be a one-to-one linear transformation.
We then know that for every $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, if $\mathbf{x} \neq \mathbf{y}$, then $T_A(\mathbf{x}) \neq T_A(\mathbf{y})$.

In particular, if $\mathbf{x} \in \mathbb{F}^n$, and $\mathbf{x} \neq \mathbf{0}_{\mathbb{F}^n}$, then $T_A(\mathbf{x}) \neq T_A(\mathbf{0}_{\mathbb{F}^n})$: by Lemma 3, $T_A(\mathbf{x}) \neq \mathbf{0}_{\mathbb{F}^m}$.
We conclude that the only vector in $\mathbb{F}^n$ that is mapped to $\mathbf{0}_{\mathbb{F}^m}$ is $\mathbf{0}_{\mathbb{F}^n}$, and so $N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\}$.

b) Let us now suppose that $N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\}$.
We will prove that for every $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, if $\mathbf{x} \neq \mathbf{y}$, then $T_A(\mathbf{x}) \neq T_A(\mathbf{y})$.
In fact, let us compare $T_A(\mathbf{x})$ and $T_A(\mathbf{y})$.
$T_A(\mathbf{x}) - T_A(\mathbf{y}) = T_A(\mathbf{x} - \mathbf{y})$, due to the linearity of T_A.
If $\mathbf{x} \neq \mathbf{y}$, then $\mathbf{x} - \mathbf{y} \neq \mathbf{0}_{\mathbb{F}^n}$. In addition, since $N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\}$, we conclude that $T_A(\mathbf{x} - \mathbf{y}) \neq \mathbf{0}_{\mathbb{F}^m}$. That is, $T_A(\mathbf{x}) - T_A(\mathbf{y}) \neq \mathbf{0}_{\mathbb{F}^m}$, or equivalently, $T_A(\mathbf{x}) \neq T_A(\mathbf{y})$.

And thus we conclude that T_A is one-to-one **iff** $N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\}$. $\blacksquare$.

Example 8

Let $A = \begin{pmatrix} 2 & 4 & 10 \\ 4 & -4 & -4 \\ 6 & 8 & 22 \end{pmatrix}$, show that $T_A : \mathbb{R}^3 \to \mathbb{R}^3$ is not one-to-one.

Solution

Let us examine the nullspace of T_A, which is the same as looking at the solution set to $A\mathbf{x} = \mathbf{0}$. We write down the augmented matrix :

$$\left(\begin{array}{ccc|c} 2 & 4 & 10 & 0 \\ 4 & -4 & -4 & 0 \\ 6 & 8 & 22 & 0 \end{array}\right).$$

Reducing it, we get:

$$\xrightarrow{R_1 \to \frac{1}{2}R_1} \left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 4 & -4 & -4 & 0 \\ 6 & 8 & 22 & 0 \end{array}\right) \xrightarrow[R_3 \to R_3 - 6R_1]{R_2 \to R_2 - 4R_1} \left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 0 & -12 & -24 & 0 \\ 0 & -4 & -8 & 0 \end{array}\right)$$

$$\xrightarrow{R_2 \to \frac{-1}{12}R_2} \left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & -4 & -8 & 0 \end{array}\right) \xrightarrow{R_3 \to R_3 + 4R_2} \left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

We see that the rank of the coefficient matrix is 2, and since $A \in M_{3\times 3}$, the number of parameters in the solution set to $A\mathbf{x} = \mathbf{0}$ is $3 - 2 = 1$.
Thus the nullspace of T_A has more vectors than only the zero vector, $N(T_A) \neq \{\mathbf{0}_{\mathbb{F}^n}\}$, and we therefore conclude that T_A is not one-to-one.

Corollary 3:

Let $A \in M_{m\times n}(\mathbb{F})$, and let T_A be the function from $\mathbb{F}^n$ to $\mathbb{F}^m$ determined by the matrix A, then T_A is one-to-one **iff**

$$N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\} \quad \textbf{iff} \quad nullity(A) = 0 \ \textbf{iff} \quad rank(A) = n.$$

Proof

We already know from Lemma 5 that $N(T_A) = N(A)$, and from Lemma 6, that T_A is one-to-one iff $N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\}$.

Thus, we immediately conclude that T_A is one-to-one iff $N(A) = \{\mathbf{0}_{\mathbb{F}^n}\}$.

Further, since $nullity(A) = n - rank(A) =$ number of parameters in $N(A)$, it follows that $N(A) = \{\mathbf{0}_{\mathbb{F}^n}\}$ iff $nullity(A) = 0$ iff $rank(A) = n$. ■

Example 9: Continuation of Example 8.

The matrix $A = \left(\begin{array}{ccc|c} 2 & 4 & 10 & 0 \\ 4 & -4 & -4 & 0 \\ 6 & 8 & 22 & 0 \end{array}\right)$ has a row echelon form $R = \left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$.

It is clear that $rank(A) = rank(R) = 2 \neq 3$, and so T_A is not one-to-one.

Alternatively, $nullity(A) = 3 - 2 = 1 \neq 0$, and so T_A is not one-to-one.

Topic 13B

Linear Transformation II

In the previous lecture, we have seen that if you have a matrix, then it is associated with a linear transformation. Moving forwards, we now suppose that our starting point is a linear transformation. We will discover that there is a matrix associated with this linear transformation, in a most natural way. This is a fundamental idea in linear algebra because this connection provides the most efficient method of actually writing the linear transformation down, and in making use of it, as we will now discover.

Reminder 1 (from Topic 11A).

In $\mathbb{F}^2$ the set of basic vectors, $S = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} = \{\mathbf{e}_1, \mathbf{e}_2\} = \{\mathbf{i}, \mathbf{j}\}$,

is very fundamental, and it is known as the standard basis in $\mathbb{F}^2$.

We also have its extension to $\mathbb{F}^n$:

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\} = \{\mathbf{e}_1, \cdots, \mathbf{e}_i, \cdots, \mathbf{e}_n\}$$

known as the standard basis in $\mathbb{F}^n$, and where $\mathbf{e}_i$ has the number 0 in each component except in the i^{th} component which is a 1.

Reminder 2

Let $T : \mathbb{F}^n \to \mathbb{F}^m$, be a **linear transformation**, this means that for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$, and for all $c \in \mathbb{F}$,

$$\begin{cases} T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) \\ T(c\,\mathbf{x}) = c\,T(\mathbf{x}) \end{cases}$$

Example 10

Let us examine the consequences of linearity in a special case when $\mathbb{F}^n = \mathbb{F}^m = \mathbb{F}^2$, that is, we have $T : \mathbb{F}^2 \to \mathbb{F}^2$. Let $\mathbf{x} \in \mathbb{F}^2$.

Suppose that $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, then

$$\begin{aligned} T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) &= T\left(\begin{pmatrix} x_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \end{pmatrix}\right) = T\left(x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\ &= x_1\, T\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + x_2\, T\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \text{ (using the linearity)} \\ &= \left(T\left(\mathbf{e}_1\right), T(\mathbf{e}_2)\right) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \left(T\left(\mathbf{e}_1\right), T(\mathbf{e}_2)\right) \mathbf{x}. \end{aligned}$$

This shows us that the actual effect of the linear transformation, T, can be replicated by the introduction of a matrix. In addition, this matrix, has columns which are constructed by applying the function, T, to the basic vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ in $\mathbb{F}^2$.

This means that, if we know what the linear transformation does to just these two (basis) vectors, then we know what it does to all vectors in $\mathbb{F}^2$. Finally, the actual value of $T(\mathbf{x})$, can be easily obtained by matrix multiplication, of the component vector $\mathbf{x}$, by this matrix $(T(\mathbf{e}_1), T(\mathbf{e}_2))$. This result extends to higher dimensions.

Definition 7: Matrix representation of a linear transformation (in the standard basis)

Let $T : \mathbb{F}^n \to \mathbb{F}^m$, be a **linear transformation.** We define the matrix representation of the linear transformation in the standard basis, standard matrix for short, $[T]_S$, to mean the $(m \times n)$ matrix whose columns are the images of the basic vectors in the standard basis in $\mathbb{F}^n$:

$$[T]_S = \left(T \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, T \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, T \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right) = (T(\mathbf{e}_1), \ldots, T(\mathbf{e}_i), \ldots, T(\mathbf{e}_n))$$

for $i = 1, \ldots, n$.

Remark 2: Brief explanation of this notation for the standard matrix.

(i) There is a T to remind us about the function that we are going to represent.

(ii) The square brackets in $[T]_S$ is our way of indicating that we are converting a linear transformation into a matrix.

(iii) The S indicates that the standard basis is being used for both the domain and the codomain. Later on in the course we will investigate the consequences of using different bases in either or both of these two sets.

Remark 3:

Let $A \in M_{m\times n}$ and T_A be the function determined by the matrix A. Then $[T_A]_S = A$.

Remark 4:

Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation with standard matrix $[T]_S$.

Let $T_{[T]_S}$ be the function determined by the matrix $[T]_S$, then $T_{[T]_S} = T$.

Remark 5:

Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation.

(a) It follows from Corollary 2 (Topic 13A) that T is onto **iff** $rank([T]_S) = m$.

(b) It follows from Corollary 3 (Topic 13A) that T is one-to-one **iff** $rank([T]_S) = n$.

Lemma 7

Let $T : \mathbb{F}^n \to \mathbb{F}^m$, be a **linear transformation**. If $\mathbf{x} \in \mathbb{F}^n$, then

$$T(\mathbf{x}) = [T]_S\,\mathbf{x}.$$

Proof

Let $\mathbf{x} \in \mathbb{F}^n$ with $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, we then have:

$$T(\mathbf{x}) = T\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = T\left(\begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{pmatrix}\right)$$

$$= T\left(x_1\,\mathbf{e}_1 + x_2\,\mathbf{e}_2 + \cdots + x_n\,\mathbf{e}_n\right)$$

$$= x_1 T\left(\mathbf{e}_1\right) + x_2 T\left(\mathbf{e}_2\right) + \cdots + x_n T\left(\mathbf{e}_n\right) \qquad \text{(using the linearity)}$$

$$= \left(T\left(\mathbf{e}_1\right), T\left(\mathbf{e}_2\right), \cdots, T\left(\mathbf{e}_n\right)\right)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = [T]_S \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}. \qquad \blacksquare$$

This is quite a lovely result.
First of all, the matrix representation provides us with a way of actually writing the linear transformation down. We may not think that this is a great accomplishment, but when we try to write down the functions in the next few examples without using linear algebra, we will be faced with a challenge perhaps unsurmountable.
Secondly, the construction of this standard matrix is very simple, it is only required to find the images of the n standard basis vectors, and build the standard matrix column by column from these images.
Thirdly, once we have this standard matrix, then it is very straightforward to find the image of any vector in $\mathbb{F}^n$ under the linear transformation. We just multiply the (component of the) vector by the matrix representation.

This is also a remarkable result, let me explain why.
In a first year calculus course, we deal with functions, $f : \mathbb{R} \to \mathbb{R}$. For almost all of them, knowing what happens at **one point** does not tell you about what happens to **other points**. For example, if we have a parabola and we are told that 1 is mapped to 5, we have no information about where the real number 2 is mapped to. And although we all know that $\sin(\frac{\pi}{6}) = \frac{1}{2}$, that does not help us if we need $\sin(1)$.
However, there is one special function encountered in first year calculus, with the special property that if we know the image of just one single (non-zero) point, then we know exactly what the function is, and thus we know the images of all points. This is a straight line through the origin. If we have a straight line through the origin, and we are told that 3 is mapped to 6, then this line must be $y = 2x$, and we know the image of any real number.
In fact, we have the result:

Lemma 8

Let $f : \mathbb{R} \to \mathbb{R}$ be a linear transformation.

If $p \in \mathbb{R} \setminus \{0\}$ with $f(p) = \alpha$, for some $\alpha \in \mathbb{R}$. Then $f(x) = \dfrac{\alpha}{p}\, x$.

Proof

Let $p \in \mathbb{R} \setminus \{0\}$ with $f(p) = \alpha$, for some $\alpha \in \mathbb{R}$. Then by linearity:

$$f(x) = f(x\left(\frac{p}{p}\right)) = f(\left(\frac{x}{p}\right)\ p) = \left(\frac{x}{p}\right)\ f(p) = \left(\frac{x}{p}\right)\ \alpha = \frac{\alpha}{p}\, x \qquad \blacksquare$$

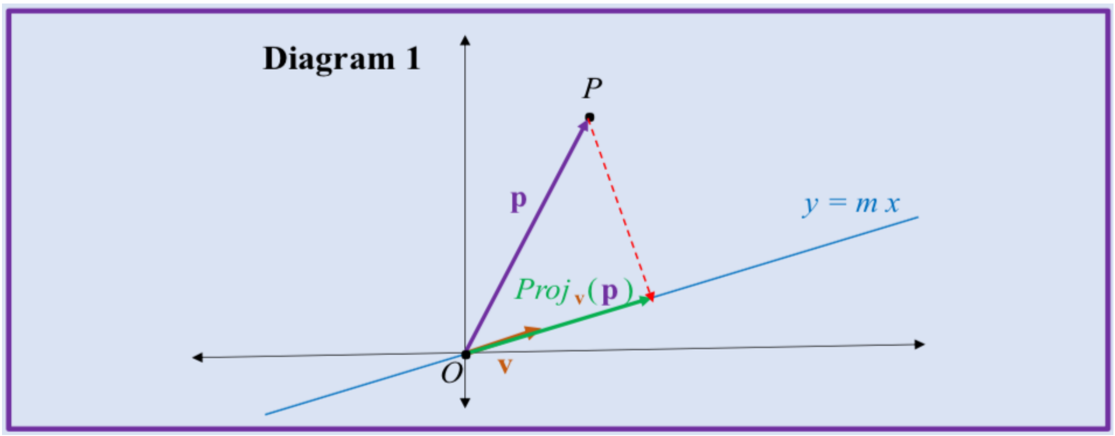
Thus the only functions from first year calculus that we will meet in this course are straight lines through the origin.

Linear transformations are just the generalizations of straight lines through the origin to higher dimensions. You only need to know the images of n-vectors in $\mathbb{F}^n$ and then you know the images of all the vectors in $\mathbb{F}^n$.

Example 11: Projection onto a line through the origin.

Consider the transformation, projection onto the line $y = mx$ in $\mathbb{R}^2$. We will refer to this function as $Proj_{\mathbf{v}}$, where $\mathbf{v}$ is a vector which is tangential to the line, so we can chose $\mathbf{v} = (1, m)^T$.

Find the standard matrix representation of $Proj_{\mathbf{v}}$.



We recall that if $\mathbf{w}$ is a vector in $\mathbb{R}^2$, the projection of $\mathbf{w}$ onto $\mathbf{v}$ is obtained by:

$$Proj_{\mathbf{v}}(\mathbf{w}) = \left(\frac{\mathbf{w} \bullet \mathbf{v}}{\mathbf{v} \bullet \mathbf{v}}\right)\mathbf{v}, \text{ so that:}$$

$$Proj_{\mathbf{v}}(\mathbf{e}_1) = \left(\frac{\begin{pmatrix} 1 \\ 0 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ m \end{pmatrix}}{\begin{pmatrix} 1 \\ m \end{pmatrix} \bullet \begin{pmatrix} 1 \\ m \end{pmatrix}}\right)\begin{pmatrix} 1 \\ m \end{pmatrix} = \frac{1}{1+m^2}\begin{pmatrix} 1 \\ m \end{pmatrix} = \begin{pmatrix} \dfrac{1}{1+m^2} \\ \dfrac{m}{1+m^2} \end{pmatrix}$$

$$Proj_{\mathbf{v}}(\mathbf{e}_2) = \left(\frac{\begin{pmatrix} 0 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ m \end{pmatrix}}{\begin{pmatrix} 1 \\ m \end{pmatrix} \bullet \begin{pmatrix} 1 \\ m \end{pmatrix}}\right)\begin{pmatrix} 1 \\ m \end{pmatrix} = \frac{m}{1+m^2}\begin{pmatrix} 1 \\ m \end{pmatrix} = \begin{pmatrix} \dfrac{m}{1+m^2} \\ \dfrac{m^2}{1+m^2} \end{pmatrix}$$

and we have

$$[Proj_{\mathbf{v}}]_S = \begin{pmatrix} \dfrac{1}{1+m^2} & \dfrac{m}{1+m^2} \\ \dfrac{m}{1+m^2} & \dfrac{m^2}{1+m^2} \end{pmatrix} = \frac{1}{1+m^2}\begin{pmatrix} 1 & m \\ m & m^2 \end{pmatrix}.$$

For instance, the image of the vector $\mathbf{w} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ under the projection onto the line $y = -2x$, is:

$$Proj_{\mathbf{v}}(\mathbf{w}) = \frac{1}{1+(-2)^2}\begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} -5 \\ 10 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

Is the linear transformation $Proj_{\mathbf{v}}$, the projection onto the line $y = mx$, an onto function? We answer this in two ways.

First of all, we think about the geometry, and the answer is no. All images lie on the line $y = mx$, and nowhere else in $\mathbb{R}^2$.

Equivalently, $R(Proj_{\mathbf{v}})$ is $Span\left(\left\{\begin{pmatrix} 1 \\ m \end{pmatrix}, \begin{pmatrix} m \\ m^2 \end{pmatrix}\right\}\right) = Span\left(\left\{\begin{pmatrix} m \\ m^2 \end{pmatrix}\right\}\right) \neq \mathbb{R}^2$.

Is $Proj_{\mathbf{v}}$, the projection onto the line $y = mx$, a one-to-one function?
We answer this in two ways.

First of all, we think about the geometry, and the answer is no. Many different points have the same image when they are projected onto the line.

Equivalently, the nullspace is obtained by solving $\dfrac{1}{1+m^2}\begin{pmatrix} 1 & m \\ m & m^2 \end{pmatrix}\mathbf{x} = \mathbf{0}$.

Since the coefficient matrix has rank of 1, then there is a parameter in the solution set, and thus the nullity is not 0: this linear transformation is therefore not one-to-one.

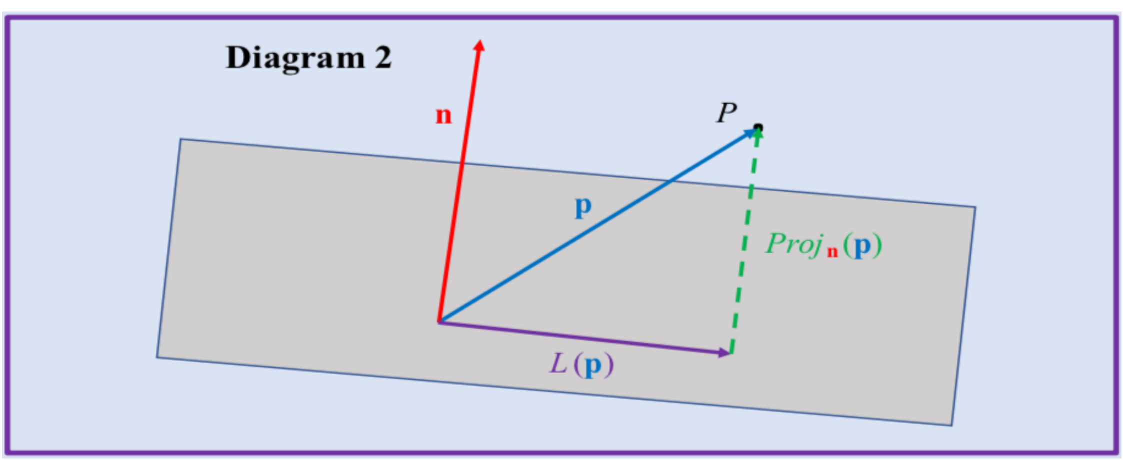
Note that if you want to project onto a line which does NOT pass through the origin, then you can project on the parallel line through the origin and then translate the solution. Also note that this function is not an example of a linear transformation, since $\mathbf{0}$ *would not be mapped to* $\mathbf{0}$.

Example 12: Projection onto a plane through the origin.

Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be the projection onto the plane $3x - 4y + 5z = 0$.

This is a linear transformation, and we will find the standard matrix of this function.

In order to obtain the projection of any vector, $\mathbf{w}$, onto the plane, we can first of all, project $\mathbf{w}$ onto a normal $\mathbf{n} = (3, -4, 5)^T$ to the plane. The difference, $\mathbf{w} - Proj_{\mathbf{n}}(\mathbf{w})$, which is the remainder, will yield the projection of $\mathbf{w}$ onto the plane.



$$Proj_{\mathbf{n}}(\mathbf{e}_1) = \frac{\mathbf{e}_1 \bullet \mathbf{n}}{\mathbf{n} \bullet \mathbf{n}} \mathbf{n} = \frac{3}{50} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \text{ so that}$$

$$L(\mathbf{e}_1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{3}{50} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} = \frac{1}{50} \begin{pmatrix} 41 \\ 12 \\ -15 \end{pmatrix}.$$

$$Proj_{\mathbf{n}}(\mathbf{e}_2) = \frac{\mathbf{e}_2 \bullet \mathbf{n}}{\mathbf{n} \bullet \mathbf{n}} \mathbf{n} = \frac{(-4)}{50} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \text{ so that}$$

$$L(\mathbf{e}_2) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \frac{(-4)}{50} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} = \frac{1}{50} \begin{pmatrix} 12 \\ 34 \\ 20 \end{pmatrix}.$$

$$Proj_{\mathbf{n}}(\mathbf{e}_3) = \frac{\mathbf{e}_3 \bullet \mathbf{n}}{\mathbf{n} \bullet \mathbf{n}} \mathbf{n} = \frac{5}{50} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \text{ so that}$$

$$L(\mathbf{e}_3) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \frac{5}{50} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} = \frac{1}{50} \begin{pmatrix} -15 \\ 20 \\ 25 \end{pmatrix}.$$

We can thus write that

$$[L]_S = \frac{1}{50} \begin{pmatrix} 41 & 12 & -15 \\ 12 & 34 & 20 \\ -15 & 20 & 25 \end{pmatrix}.$$

For instance, what are the projections of $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}$ and $\begin{pmatrix} 10 \\ 15 \\ 6 \end{pmatrix}$ onto the plane?

Using $[L]_S$ we get quickly that:

$$L\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}\right) = \frac{1}{50} \begin{pmatrix} 41 & 12 & -15 \\ 12 & 34 & 20 \\ -15 & 20 & 25 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{50} \begin{pmatrix} 20 \\ 140 \\ 100 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 2 \\ 14 \\ 10 \end{pmatrix},$$

$$L\left(\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}\right) = \frac{1}{50} \begin{pmatrix} 41 & 12 & -15 \\ 12 & 34 & 20 \\ -15 & 20 & 25 \end{pmatrix} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} = \frac{1}{50} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$L\left(\begin{pmatrix}10\\15\\6\end{pmatrix}\right)=\frac{1}{50}\begin{pmatrix}41&12&-15\\12&34&20\\-15&20&25\end{pmatrix}\begin{pmatrix}10\\15\\6\end{pmatrix}=\frac{1}{50}\begin{pmatrix}500\\750\\300\end{pmatrix}=\begin{pmatrix}10\\15\\6\end{pmatrix}.$$

The last two answers should not be a surprise.

The vector $\begin{pmatrix}3\\-4\\5\end{pmatrix}$ is normal to the plane, and thus it has zero projection on the plane.

The vector $\begin{pmatrix}10\\15\\6\end{pmatrix}$ is a vector which lies in the plane, and thus its projection is itself.

Is this function onto? We answer this in two ways.

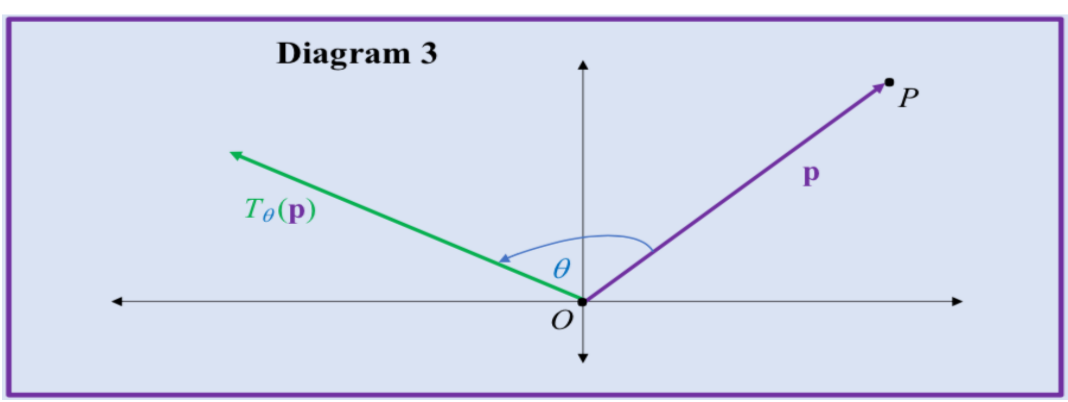
First of all, we think about the geometry, and the answer is no. All images lie on the plane $3x-4y+5z=0$, and nowhere else in $\mathbb{R}^3$.
Equivalently, we could investigate the range and show that it is not $\mathbb{R}^3$, we will return to this type of calculation later on.

Is this function one-to-one? We answer this in two ways.

First of all, we think about the geometry, and the answer is no. Many different points have the same image when they are projected onto the plane.
Equivalently, the nullspace is obtained by solving $[L]_S\,\mathbf{x}=\mathbf{0}$.
Since the coefficient matrix has rank of 2, (check it) then there is a parameter in the solution set, and we conclude that this linear transformation is not one-to-one.

Example 13: Rotation about the origin by an angle θ.

Consider the linear transformation, rotation about the origin by an angle of θ anticlockwise in $\mathbb{R}^2$, which we denote by $T_\theta:\mathbb{R}^2\to\mathbb{R}^2$. This is illustrated in Diagram 3 below. Find the standard matrix representation of this function.



We perform a little elementary trigonometry to get:

$$T_\theta(\mathbf{e}_1) = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}, \quad T_\theta(\mathbf{e}_2) = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix},$$

so that the matrix representation is

$$[T_\theta]_S = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Suppose we want the image of the vector $\begin{pmatrix} 2 \\ -3 \end{pmatrix}$, when $\theta = \frac{\pi}{3}$, then we have:

$$\begin{aligned} T_{\frac{\pi}{3}}\left(\begin{pmatrix} 2 \\ -3 \end{pmatrix}\right) &= \begin{pmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & \cos(\frac{\pi}{3}) \end{pmatrix}\begin{pmatrix} 2 \\ -3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & \frac{-\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 2 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 + \frac{3\sqrt{3}}{2} \\ \sqrt{3} - \frac{3}{2} \end{pmatrix}. \end{aligned}$$

Note that values of trigonometric functions of special angles should be known.

Is this function onto? We answer this in two ways.

First of all, we think about the geometry, and the answer is yes. Any vector in $\mathbb{R}^2$ can be obtained by a rotation from an appropriate starting vector.

Equivalently, the range is $Span\left(\left\{\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}, \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}\right\}\right)$ and it may be shown that this is $\mathbb{R}^2$.

Is this function one-to-one? We answer this in two ways.

First of all, we think about the geometry, and the answer is yes. Any two different vector have different images when they are rotated.

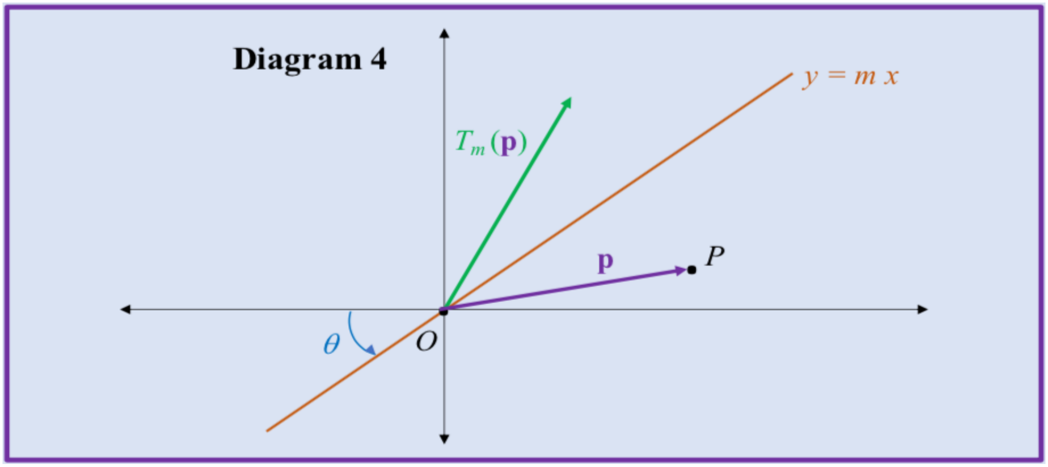
Equivalently, the nullspace is obtained by solving $\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}\mathbf{x} = \mathbf{0}$.

Since the coefficient matrix has rank of 2, then the only solution is the trivial solution and so this linear transformation is one-to-one.

Example 14: Reflection about a line through the origin.

Consider the linear transformation, reflection in the line $y = mx$, in $\mathbb{R}^2$, where $m \in \mathbb{R}$, is a constant and $m = \tan(\theta)$ for some angle θ, which we denote by T_m (see Diagram 4).

Find the standard matrix representation of this function.



We notice that the vector $\mathbf{e}_1$ is rotated **anticlockwise** by an angle 2θ, and so we know from Example 13 that

$$T_m\left(\mathbf{e}_1\right) = \begin{pmatrix} \cos(2\theta) \\ \sin(2\theta) \end{pmatrix}.$$

On the other hand, the vector $\mathbf{e}_2$ is rotated **clockwise** by an angle $2(\frac{\pi}{2} - \theta)$, that is, **anticlockwise** by an angle of $-2(\frac{\pi}{2} - \theta)$, so we know from Example 13 that

$$T_m\left(\mathbf{e}_2\right) = \begin{pmatrix} -\sin(2\theta - \pi) \\ \cos(2\theta - \pi) \end{pmatrix} = \begin{pmatrix} \sin(2\theta) \\ -\cos(2\theta) \end{pmatrix}.$$

Thus the matrix representation of T_m in the standard basis is:

$$[T_m]_S = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}.$$

It may be shown with a little bit of trigonometry that:

$$\cos(2\theta) = \frac{1 - m^2}{1 + m^2} \qquad \text{and} \qquad \sin(2\theta) = \frac{2m}{1 + m^2}.$$

We thus conclude that:

$$[T_m]_S = \frac{1}{1 + m^2} \begin{pmatrix} 1 - m^2 & 2m \\ 2m & -(1 - m^2) \end{pmatrix}.$$

For instance, what are the images of the points $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -3 \\ 2 \end{pmatrix}$ when they are reflected in the line $y = 4x$?

In this case, the matrix representation is:

$$[T_4]_S = \frac{1}{1+4^2}\begin{pmatrix} 1-4^2 & 2(4) \\ 2(4) & -(1-4^2) \end{pmatrix} = \frac{1}{17}\begin{pmatrix} -15 & 8 \\ 8 & 15 \end{pmatrix},$$

and so the vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is mapped to $[T_4]_S \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, that is,

$$T_4\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) = \frac{1}{17}\begin{pmatrix} -15 & 8 \\ 8 & 15 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{17}\begin{pmatrix} 1 \\ 38 \end{pmatrix}.$$

And the vector $\begin{pmatrix} -3 \\ 2 \end{pmatrix}$ is mapped to $[T_4]_S \begin{pmatrix} -3 \\ 2 \end{pmatrix}$, that is,

$$T_4\left(\begin{pmatrix} -3 \\ 2 \end{pmatrix}\right) = \frac{1}{17}\begin{pmatrix} -15 & 8 \\ 8 & 15 \end{pmatrix}\begin{pmatrix} -3 \\ 2 \end{pmatrix} = \frac{1}{17}\begin{pmatrix} 61 \\ 6 \end{pmatrix}.$$

Which vector is mapped to the vector $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$?

If $T_4\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$, then we have to solve:

$$T_4\left(\begin{pmatrix} a \\ b \end{pmatrix}\right) = \frac{1}{17}\begin{pmatrix} -15 & 8 \\ 8 & 15 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix},$$

and we have the system of equations

$$\begin{aligned} -15a + 8b &= 34 \\ 8a + 15b &= 51 \end{aligned}$$

This has an augmented matrix of

$$\begin{pmatrix} -15 & 8 & | \, 34 \\ 8 & 15 & | \, 51 \end{pmatrix},$$

which we row reduce in REF to

$$\begin{pmatrix} 1 & \frac{-8}{15} & | \, \frac{-34}{15} \\ 0 & \frac{289}{15} & | \, \frac{1037}{15} \end{pmatrix} \text{ and in RREF to } \begin{pmatrix} 1 & 0 & | \, \frac{-6}{17} \\ 0 & 1 & | \, \frac{61}{17} \end{pmatrix},$$

and we conclude that the vector $\begin{pmatrix} \frac{-6}{17} \\ \frac{61}{17} \end{pmatrix}$ is mapped to $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

Is this function onto? We answer this in two ways.

First of all, we think about the geometry, and the answer is yes. Any vector in $\mathbb{R}^2$ can be obtained by a reflection from an appropriate starting vector.

Equivalently, the range is $Span\left(\left\{\left(\begin{array}{c} 1-m^2 \\ 2m \end{array}\right), \left(\begin{array}{c} 2m \\ -(1-m^2) \end{array}\right)\right\}\right)$ and this is $\mathbb{R}^2$.

Is this function one-to-one? We answer this in two ways.

First of all, we think about the geometry, and the answer is yes. Any two different vectors have different images when they are reflected.

Equivalently, the nullspace is obtained by solving $\left(\begin{array}{cc} 1-m^2 & 2m \\ 2m & -(1-m^2) \end{array}\right)\mathbf{x} = \mathbf{0}$.

Since the coefficient matrix has rank of 2, then the only solution is the trivial solution and so this linear transformation is one-to-one.

Topic 13C

Linear Transformation III

Composition and Invertibility

In this topic, we will consider the composition of linear transformations and show how this links nicely with matrix multiplication. We will define the inverse of a matrix and relate this to the invertibility of the function determined by the matrix.

Definition 8: Composition of functions

Let $T_1 : \mathbb{F}^n \to \mathbb{F}^m$ and $T_2 : \mathbb{F}^m \to \mathbb{F}^p$ be functions.

We define the function $T = T_2 \circ T_1$, $T : \mathbb{F}^n \to \mathbb{F}^p$, by

$$T(\mathbf{x}) = (T_2 \circ T_1)(\mathbf{x}) = T_2(T_1(\mathbf{x})).$$

T is called the **composite function** of T_2 and T_1.

Lemma 9: Composition of linear transformation is linear.

Let $T_1 : \mathbb{F}^n \to \mathbb{F}^m$ and $T_2 : \mathbb{F}^m \to \mathbb{F}^p$.

If T_1 and T_2 are both linear transformations, then the composite function of T_2 and T_1 is also a linear transformation.

Proof

Since T_1 is linear, we have:

$$T_1(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1T_1(\mathbf{x}_1) + c_2T_1(\mathbf{x}_2), \text{ for all } \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{F}^n, \ c_1, c_2 \in \mathbb{F}.$$

Since T_2 is linear we have:

$$T_2(d_1\mathbf{y}_1 + d_2\mathbf{y}_2) = d_1T_2(\mathbf{y}_1) + d_2T_2(\mathbf{y}_2), \text{ for all } \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{F}^m, \ d_1, d_2 \in \mathbb{F}.$$

Consider $(T_2 \circ T_1)(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = T_2\left(c_1T_1(\mathbf{x}_1) + c_2T_1(\mathbf{x}_2)\right)$.

We can let $T_1(\mathbf{x}_1) = \mathbf{y}_1$, and $T_1(\mathbf{x}_2) = \mathbf{y}_2$, so that:

$$\begin{aligned}(T_2 \circ T_1)(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = T_2(c_1\mathbf{y}_1 + c_2\mathbf{y}_2) &= c_1T_2(\mathbf{y}_1) + c_2T_2(\mathbf{y}_2), \quad \text{by linearity of } T_2 \\ &= c_1T_2(T_1(\mathbf{x}_1)) + c_2T_2(T_1(\mathbf{x}_2))\end{aligned}$$

and we conclude that $(T_2 \circ T_1)(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1(T_2 \circ T_1)(\mathbf{x}_1) + c_2(T_2 \circ T_1)(\mathbf{x}_2)$,

i.e. $(T_2 \circ T_1)$ is linear. $\blacksquare$

Lemma 10: The matrix of the composite function.

Let $T_1 : \mathbb{F}^n \to \mathbb{F}^m$ and $T_2 : \mathbb{F}^m \to \mathbb{F}^p$ be linear transformations.

Let the composite function of T_2 and T_1 be T, that is, let $T = T_2 \circ T_1$. Then

$$[T]_S = [T_2 \circ T_1]_S = [T_2]_S\,[T_1]_S\,.$$

Proof

Let the standard basis for $\mathbb{F}^n$ be $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ and the one for $\mathbb{F}^m$ be $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}$.

Suppose that we let:

$$A = [T_1]_S\,, \text{ so that } a_{kj} = (T_1(\mathbf{e}_j))_k \text{ and}$$

$$B = [T_2]_S\,, \text{ so that } b_{ik} = (T_2(\mathbf{f}_k))_i\,.$$

In order to obtain $[T]_S$, we must apply T to each of the vectors in $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$,

we then have:

$$\begin{aligned}
([T]_S)_{ij} &= (T(\mathbf{e}_j))_i && \text{by definition of the matrix representation}\\
&= ((T_2 \circ T_1)(\mathbf{e}_j))_i = \left(T_2\left(\sum_{k=1}^{m} a_{kj}\mathbf{f}_k\right)\right)_i && \text{by definition of } a_{kj}\\
&= \left(\sum_{k=1}^{m} a_{kj} T_2(\mathbf{f}_k)\right)_i && \text{using the linearity of } T_2\\
&= \sum_{k=1}^{m} a_{kj}\,(T_2(\mathbf{f}_k))_i && \text{using the linearity of taking a component}\\
&= \sum_{k=1}^{m} a_{kj} b_{ik} = \sum_{k=1}^{m} b_{ik} a_{kj} = (BA)_{ij} && \text{by the definition of matrix multiplication}\\
&= ([T_2]_S\,[T_1]_S)_{ij}\,.
\end{aligned}$$

Since all their entries are equal for all $i = 1, \dots, p$ and $j = 1, \dots, n$, we conclude that the two matrices are identical, and thus $[T]_S = [T_2]_S\,[T_1]_S$. ■

This gives us a very efficient way of obtaining the matrix representation of a composite function. It also explains the reasons behind the definition of matrix multiplication.

Example 15

Let $A = \begin{pmatrix} 1+i & -3i & -2 \\ 2 & 1+2i & 4i \end{pmatrix}$ be a matrix in $M_{2\times 3}(\mathbb{C})$ and

$B = \begin{pmatrix} 3i & 1+i \\ 1 & 0 \\ 1-i & 3-2i \end{pmatrix}$ be a matrix in $M_{3\times 2}(\mathbb{C})$.

Let T_A and T_B be the functions determined by A and B, respectively.
What is the standard matrix of their composite function, $T_B \circ T_A$?

Solution

$[T_B \circ T_A]_S = [T_B]_S\,[T_A]_S = BA$

$$= \begin{pmatrix} 3i & 1+i \\ 1 & 0 \\ 1-i & 3-2i \end{pmatrix} \begin{pmatrix} 1+i & -3i & -2 \\ 2 & 1+2i & 4i \end{pmatrix} = \begin{pmatrix} 5i-1 & 8+3i & -4-2i \\ 1+i & -3i & -2 \\ 8-4i & 4+i & 6+14i \end{pmatrix}.$$

Example 16

Find the standard matrix for the linear transformation, T, on $\mathbb{R}^2$ which is defined as, T_1, a rotation anticlockwise about the origin by an angle of $\frac{\pi}{3}$ radians, followed by T_2, a projection onto the line $y = -3x$.

Solution

Example 13 (Topic 13B) showed that $[T_1]_S = \begin{pmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & \cos(\frac{\pi}{3}) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}$.

Example 11 (Topic 13B) showed that $[T_2]_S = \dfrac{1}{1+(-3)^2}\begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix} = \dfrac{1}{10}\begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix}$.

Thus we have:

$$[T]_S = [T_2]_S\,[T_1]_S = \frac{1}{10}\begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix}\begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 1-3\sqrt{3} & -3-\sqrt{3} \\ -3+9\sqrt{3} & 3\sqrt{3}+9 \end{pmatrix}.$$

A special case of composition arises when $T : \mathbb{F}^n \to \mathbb{F}^n$, in which case we can apply the function T more than one time.

Definition 9: T^p

Let $T : \mathbb{F}^n \to \mathbb{F}^n$ and let $p > 1$ be an integer. We then define T^p inductively by:

$$T^p = T \circ T^{p-1}.$$

Note that $T^0 = T_I$, the identity transformation, defined by $T_I(\mathbf{x}) = \mathbf{x}, \forall \mathbf{x} \in \mathbb{F}^n$.

Corollary 4: (of Lemma 10)

Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation and let $p > 1$ be an integer, then

$$[T^p]_S = ([T]_S)^p \,.$$

Proof: Direct application of Lemma 10 and induction.

Example 17

Consider the linear transformation, rotation about the origin by an angle of θ anticlockwise in the plane, which we denote by $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$.

In Example 13 (Topic 13B) we find that $[T_\theta]_S = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$.

Now suppose we have the linear transformation $L = T_\theta \circ T_\theta$, then we have:

$$
\begin{aligned}
[L]_S = ([T_\theta]_S)^2 &= \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \\
&= \begin{pmatrix} \cos^2(\theta) - \sin^2(\theta) & -2\cos(\theta)\sin(\theta) \\ 2\cos(\theta)\sin(\theta) & \cos^2(\theta) - \sin^2(\theta) \end{pmatrix} \\
&= \begin{pmatrix} \cos(2\theta) & -\sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{pmatrix},
\end{aligned}
$$

which is, of course, the standard matrix for a rotation of 2θ anticlockwise.

NOTE: the next section is defined ONLY for square matrices.

Invertibility of Matrices

Definition 10: Invertibility of a matrix.

Let $A \in M_{n\times n}$. We say that A is **invertible** to mean that there exists $B \in M_{n\times n}$ such that $AB = BA = I_n$.

The matrix B is called **an inverse** of the matrix A. If follows from this definition that B is invertible, and the matrix A is called an inverse of the matrix B. We also say that A and B are inverses of each other.

Definition 11: Singularity of a matrix.

We say that A is **singular** to mean that A is **not** invertible.

An invertible matrix is also referred to as a non-singular matrix.

Examples 18

- Show that $B = \begin{pmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ \frac{7}{2} & \frac{3}{2} & \frac{1}{2} \end{pmatrix}$ is an inverse of the matrix $A = \begin{pmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{pmatrix}$.

Solution

$$\begin{pmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ \frac{7}{2} & \frac{3}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \textbf{ and}$$

$$\begin{pmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{pmatrix} \begin{pmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ \frac{7}{2} & \frac{3}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We conclude that B is an inverse of A.

- Show that $D = \begin{pmatrix} 1+i & -2+3i \\ 7-5i & 4+6i \end{pmatrix}$ is an inverse of $C = \dfrac{-3+21i}{450} \begin{pmatrix} 4+6i & 2-3i \\ -7+5i & 1+i \end{pmatrix}$.

Solution

$$\begin{aligned} DC &= \begin{pmatrix} 1+i & -2+3i \\ 7-5i & 4+6i \end{pmatrix} \left(\frac{-3+21i}{450}\right) \begin{pmatrix} 4+6i & 2-3i \\ -7+5i & 1+i \end{pmatrix} \\ &= \frac{-3+21i}{450} \begin{pmatrix} 1+i & -2+3i \\ 7-5i & 4+6i \end{pmatrix} \begin{pmatrix} 4+6i & 2-3i \\ -7+5i & 1+i \end{pmatrix} \\ &= \frac{-3+21i}{450} \begin{pmatrix} -3-21i & 0 \\ 0 & -3-21i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} CD &= \frac{-3+21i}{450} \begin{pmatrix} 4+6i & 2-3i \\ -7+5i & 1+i \end{pmatrix} \begin{pmatrix} 1+i & -2+3i \\ 7-5i & 4+6i \end{pmatrix} \\ &= \frac{-3+21i}{450} \begin{pmatrix} -3-21i & 0 \\ 0 & -3-21i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

We conclude that D is an inverse of C and vice-versa.

- Show that the matrix $F = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$ is singular.

Solution

Let $G{=}\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and consider the product:

$$FG = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+2c & b+2d \\ 0 & 0 \end{pmatrix}.$$

Independently of the choices of the scalars a, b, c, d, it is impossible to make FG equal to I_2. We conclude that F is singular.

We will show how to obtain inverses in the next topic.

Lemma 11: Unique inverse

Let $A \in M_{n\times n}(\mathbb{F})$.

If A is **invertible** (that is, there exists a matrix B such that $AB = BA = I_n$), then the matrix B is unique.

Proof

Let $A \in M_{n\times n}(\mathbb{F})$ be invertible, i.e., there exists a matrix B such that $AB = BA = I_n$.

Suppose that there exists another matrix $C \in M_{n\times n}$ such that $AC = CA = I_n$.
Then we can write:

$$B = B\left(I_n\right) = B\left(AC\right) = BAC = \left(BA\right)C = I_nC = C.$$

We conclude that B is unique. $\blacksquare$

We can now refer to **the** inverse of A, and we denote it by A^{-1}. (In a similar way that the scalar 2 is invertible and we denote its (multiplicative) inverse by 2^{-1}).

The matrices A and A^{-1} are inverses of each other, so, in particular:

$$\left(A^{-1}\right)^{-1} = A.$$

Lemma 12

If $A \in M_{n\times n}(\mathbb{F})$ is invertible, then

$$A\mathbf{x} = \mathbf{b} \textbf{ has a unique solution, } \mathbf{z} = A^{-1}\mathbf{b}\,, \text{ for all } \mathbf{b} \in \mathbb{F}^n.$$

Proof

Let A be invertible. Then the inverse matrix A^{-1} exists.
Let $\mathbf{z} = A^{-1}\mathbf{b}$, where $\mathbf{b} \in \mathbb{F}^n$.

Since $A(\mathbf{z}) = A(A^{-1}\mathbf{b}) = AA^{-1}(\mathbf{b}) = I_n\mathbf{b} = \mathbf{b}$, then $\mathbf{z}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

If $\mathbf{y}$ is a solution of $A\mathbf{x} = \mathbf{b}$, then $A\mathbf{y} = \mathbf{b}$, and thus $A^{-1}(A\mathbf{y}) = A^{-1}\mathbf{b} = \mathbf{z}$,
and we conclude that $I_n\,\mathbf{y} = \mathbf{z}$, that is, $\mathbf{y} = \mathbf{z}$. ■.

We will show that this result is an iff in the Corollary 5 (of Lemma 16).

Lemma 13: Properties of the inverse.

Let A and B be invertible $M_{n\times n}(\mathbb{F})$ matrices and let c be a non-zero scalar in $\mathbb{F}$, then:

(i) A^T is invertible and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

(ii) cA is invertible and $(cA)^{-1} = c^{-1}A^{-1}$.

(iii) AB is invertible and $(AB)^{-1} = \left(B^{-1}\right)\left(A^{-1}\right)$,

that is, the inverse of the product is the product of the inverses IN REVERSE,
when both matrices are invertible.

(iv) If $C, D \in M_{n\times p}(\mathbb{F})$ and $AC = AD$, then $C = D$.

(v) If $C \in M_{n\times p}(\mathbb{F})$ and $AC = \mathbb{O}_{n\times p}$, then $C = \mathbb{O}_{n\times p}$.

Lemma 14: Inverses of elementary matrices.

Any elementary matrix is invertible, and the inverse of an elementary matrix
is an elementary matrix and it has the same type as the original elementary matrix.

In fact we can say more:

I) The inverse of a type I elementary matrix is itself.

II) If E is an elementary matrix obtained by scaling row i of I_n by the non-zero
scalar m, then E^{-1} is the elementary matrix obtained by scaling row i of I_n by m^{-1}.

III) If E is an elementary matrix obtained by adding m times row j of I_n to row i, then
E^{-1} is the elementary matrix obtained by subtracting m times row j of I_n from row i.

Example 19

$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ is a type I elementary matrix $(R_1 \leftrightarrow R_3)$, its inverse is $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

$\begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$ is a type II elementary matrix $(R_1 \to 5R_1)$, its inverse is $\begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \end{pmatrix}$.

$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$ is a type III elementary matrix $(R_2 \to R_2 - 2R_3)$, its inverse is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$.

We now consider invertibility of functions.

Definition 12

Let $T_1 : \mathbb{F}^n \to \mathbb{F}^m$ be a function, then we say that T_1 is **invertible** to mean that there exists another function $T_2 : \mathbb{F}^m \to \mathbb{F}^n$, such that

$$T_2 \circ T_1 = T_{I_{\mathbb{F}^n}}, \quad \text{the identity function on } \mathbb{F}^n,$$

and

$$T_1 \circ T_2 = T_{I_{\mathbb{F}^m}}, \text{ the identity function on } \mathbb{F}^m.$$

We recall from previous courses (e.g. MATH 135 and MATH 137) that a **function** is invertible **iff** the function is both onto and one-to-one.

We also recall from Remark 5 (Topic 13B) that if a **linear transformation** from $\mathbb{F}^m$ to $\mathbb{F}^n$ is both onto and one-to-one, then $m = n$.

Combining these two results, we conclude that only linear transformations which have the same domain and codomain, $\mathbb{F}^n$, have the **possibility** of being invertible.

Suppose we have $T_1 : \mathbb{F}^n \to \mathbb{F}^n$, a linear transformation, then T_1 is invertible means that there exists another function $T_2 : \mathbb{F}^n \to \mathbb{F}^n$, such that

$$T_2 \circ T_1 = T_1 \circ T_2 = T_{I_{\mathbb{F}^n}}, \text{ the identity function on } \mathbb{F}^n.$$

Lemma 15

Let T be a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^n$.

If T is invertible, then its inverse is unique and linear.

Proof

The proof of the inverse, say S, being unique is almost identical to that of the inverse of a matrix being unique, when these inverses exist.

We prove that S is linear. Let $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{F}^n$, and $c_1, c_2 \in \mathbb{F}$.

Note that since T is onto, then there exist $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{F}^n$ such that

$$T(\mathbf{x}_1) = \mathbf{y}_1 \text{ and } T(\mathbf{x}_2) = \mathbf{y}_2.$$

Note also that $S(T(\mathbf{x}_1)) = S(\mathbf{y}_1) = \mathbf{x}_1$, and similarly $S(T(\mathbf{x}_2)) = S(\mathbf{y}_2) = \mathbf{x}_2$.

We then have:

$$(S \circ T)(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = T_{I_{\mathbb{F}^n}}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\mathbf{x}_1 + c_2\mathbf{x}_2,$$

and since T is linear, we also have for all $c_1, c_2 \in \mathbb{F}^n$:

$$(S \circ T)(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = S(T(c_1\mathbf{x}_1 + c_2\mathbf{x}_2)) = S(c_1T(\mathbf{x}_1) + c_2T(\mathbf{x}_2)) = S(c_1\mathbf{y}_1 + c_2\mathbf{y}_2),$$

and thus : $(S \circ T)(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = S(c_1\mathbf{y}_1 + c_2\mathbf{y}_2) = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = c_1S(\mathbf{y}_1) + c_2S(\mathbf{y}_2),$

yielding that : $S(c_1\mathbf{y}_1 + c_2\mathbf{y}_2) = c_1S(\mathbf{y}_1) + c_2S(\mathbf{y}_2).$

We thus conclude that S is linear. ■

Since the inverse of a linear transformation, T, if it exists, is unique, then we refer to it as **the inverse** and denote it by T^{-1}.

Lemma 16

Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation.

T is invertible **iff** $[T]_S$ is an invertible matrix. In this case, $[T^{-1}]_S = \left([T]_S\right)^{-1}$.

This provides us with a very simple way of writing down the inverse of an invertible linear transformation once we have a way of obtaining the inverse of a matrix.

Proof

Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation.

Suppose T is invertible. Then we have:

$$T^{-1} \circ T = T_{I_n} \text{ and } T \circ T^{-1} = T_{I_n}.$$

Taking the matrix representation and applying Lemma 10 gives:

$$[T^{-1} \circ T]_S = [T^{-1}]_S\,[T]_S = [T_I]_S = I_n \quad \text{and} \quad [T \circ T^{-1}]_S = [T]_S\,[T^{-1}]_S = [T_I]_S = I_n.$$

We conclude that $[T]_S$ is invertible and its inverse matrix is $[T^{-1}]_S$.

Suppose now that $[T]_S$ is an invertible matrix A and define $L : \mathbb{F}^n \to \mathbb{F}^n$ by $L = T_{A^{-1}}$.

We then have:

$$(L \circ T)(\mathbf{x}) = L(A\mathbf{x}) = T_{A^{-1}}(A\mathbf{x}) = A^{-1}A\mathbf{x} = \mathbf{x}, \quad \text{for all } \mathbf{x} \in \mathbb{F}^n, \quad \text{and}$$
$$(T \circ L)(\mathbf{x}) = T(T_{A^{-1}}(\mathbf{x})) = T(A^{-1}\mathbf{x}) = AA^{-1}\mathbf{x} = \mathbf{x}, \quad \text{for all } \mathbf{x} \in \mathbb{F}^n.$$

We conclude that $L \circ T = T_I$ and $T \circ L = T_I$, so that L is the inverse transformation to T, that is, T is an invertible transformation.

Thus, T is invertible iff $[T]_S$ is invertible. ■

Corollary 5: (of Lemma 16)

Let $A \in M_{n\times n}(\mathbb{F})$.

If $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b} \in \mathbb{F}^n$, then A is an invertible matrix.

Proof

Consider T_A the linear transformation defined by the matrix A.

Suppose $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b} \in \mathbb{F}^n$. We then have:

(I) $A\mathbf{x} = \mathbf{0}$ has a unique solution, $\mathbf{x} = \mathbf{0}$. Thus, $N(T_A) = \{\mathbf{0}_{\mathbb{F}^n}\}$ and so T is one-to-one.

(II) $R(T_A) = \mathbb{F}^n$, since $A\mathbf{x} = \mathbf{b}$ can be solved for any $\mathbf{b} \in \mathbb{F}^n$, and so T is onto.

We then conclude that T_A is invertible.

By Lemma 16, $[T_A] = A$ is an invertible matrix. ■

Examples 20

- Let T_A be the function determined by the matrix $A = \begin{pmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{pmatrix} = [T_A]_S$.

Show that T_A is an invertible function and write down its inverse.

Solution

We know from Example 18 that A is invertible with inverse $A^{-1} = \begin{pmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ \frac{7}{2} & \frac{3}{2} & \frac{1}{2} \end{pmatrix}$.

We conclude that T_A is an invertible function and

$$\left[T_A^{-1}\right]_S = [T_A]_S^{-1} = A^{-1}, \text{ so}$$

$$T_A^{-1}(\mathbf{x}) = \begin{pmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ \frac{7}{2} & \frac{3}{2} & \frac{1}{2} \end{pmatrix} \mathbf{x}, \text{ for all } \mathbf{x} \in \mathbb{R}^3.$$

- Let T_C be the function determined by the matrix $C = \dfrac{-3+21i}{450} \begin{pmatrix} 4+6i & 2-3i \\ -7+5i & 1+i \end{pmatrix}$.
Show that T_C is an invertible function and write down its inverse.

Solution

We know from Examples 18 that C is invertible with inverse $C^{-1} = \begin{pmatrix} 1+i & -2+3i \\ 7-5i & 4+6i \end{pmatrix}$.

We conclude that T_C is an invertible function and

$$\left[T_C^{-1}\right]_S = [T_C]_S^{-1} = C^{-1}, \text{ and so}$$

$$T_C^{-1}(\mathbf{z}) = \begin{pmatrix} 1+i & -2+3i \\ 7-5i & 4+6i \end{pmatrix} \mathbf{z}, \text{ for all } \mathbf{z} \in \mathbb{C}^2.$$

- Consider the linear transformation, the rotation about the origin by an angle of θ anticlockwise in the plane, which we denote by $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$.

Show that T_θ is invertible and write down its inverse.

We know from Example 13 (Topic 13B) that $[T_\theta]_S = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$.

If we think about it, we know that the inverse function to this linear transformation is, rotation about the origin by an angle of $-\theta$ anticlockwise in the plane, whose standard matrix is:

$$[T_{-\theta}]_S = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Let us check: the product of $[T_\theta]$ and $[T_{-\theta}]_S$ should be the identity matrix, if these two transformations are inverses of each other.

$$[T_\theta]_S\,[T_{-\theta}]_S = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$= \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 \\ 0 & \cos^2(\theta) + \sin^2(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

$$[T_{-\theta}]_S\,[T_\theta]_S = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$= \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 \\ 0 & \cos^2(\theta) + \sin^2(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

And thus, since $[T_\theta]_S\,[T_{-\theta}]_S = [T_{-\theta}]_S\,[T_\theta]_S = I_2$, then $[T_\theta]$ is an invertible matrix. We conclude that T_θ is an invertible function, with $(T_\theta)^{-1} = T_{-\theta}$.

We also have $(T_\theta)^{-1}(\mathbf{x}) = T_{-\theta}(\mathbf{x}) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \mathbf{x}$.

Definition 13: Isomorphism

An invertible linear transformation is called an **isomorphism**.

Example 21

The linear transformations T_A and T_C in Examples 20 are both isomorphisms.

$T_0 : \mathbb{C}^3 \to \mathbb{C}^3$ given by $T_0(\mathbf{z}) = \mathbf{0}$ for all $\mathbf{z} \in \mathbb{C}^3$ is not an isomorphism.

$T_I : \mathbb{C}^3 \to \mathbb{C}^3$ given by $T_I(\mathbf{z}) = \mathbf{z}$ for all $\mathbf{z} \in \mathbb{C}^3$ is an isomorphism.

Remark 6

We recall from Remark 5 (Topic 13B) that if $T : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation with $rank([T]_S) = r$, then

(a) T onto **iff** $r = m$

(b) T is one-to-one **iff** $r = n$

Now suppose that $n = m$. Then we get:

If T is one-to-one, then $n = r(= m)$, so T is onto.

If T is onto, then $m = r(= n)$, so T is one-to-one.

Thus, the linear transformation $T : \mathbb{F}^n \to \mathbb{F}^n$ is one-to-one **iff** it is onto.

When we are checking to determine whether $T : \mathbb{F}^n \to \mathbb{F}^n$ is invertible or not, we need only to check to see it is either one-to-one or onto. We do not have to check for both of these.

If $T : \mathbb{F}^n \to \mathbb{F}^n$ is a linear transformation and if T is either one-to-one or onto, then it is **both** one-to-one and onto, and thus it is an isomorphism.

Topic 14

Matrix Inverse

Lemma 1

Suppose that $A \in M_{n\times n}(\mathbb{F})$.

If there exists $B \in M_{n\times n}(\mathbb{F})$ such that $AB = I_n$, then A is invertible.

Proof

Let $A \in M_{n\times n}(\mathbb{F})$ and let B be a matrix in $M_{n\times n}(\mathbb{F})$ such that $AB = I_n$.

We first show that the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{F}^n$.

Since $AB = I_n$, then $(AB)\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$, that is, $A(B\mathbf{b}) = \mathbf{b}$, or $A\mathbf{x} = \mathbf{b}$ with $\mathbf{x} = B\mathbf{b}$.
Thus the system $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{F}^n$.

Let $T_A : \mathbb{F}^n \to \mathbb{F}^n$ be the linear transformation determined by the matrix A.
We conclude that $Col(A) = \mathbb{F}^n$, so that $R(T_A) = \mathbb{F}^n$, and thus T_A is onto.
Finally, from the Remark 6 (T13C), T_A is invertible, and so is A (Lemma 16 - T13C). ∎

This will save us quite a lot of arithmetic when we are checking our candidate inverses: we need only verify that $AB = I$, and not perform the additional check of $BA = I$.

Lemma 2: Invertibility of a matrix.

Suppose that $A \in M_{n\times n}(\mathbb{F})$, then A is invertible **iff** $rank(A) = n$.

Proof

Let $A \in M_{n\times n}(\mathbb{F})$.
Let T_A be the linear transformation determined by the matrix A ($T_A : \mathbb{F}^n \to \mathbb{F}^n$).

A is invertible **iff** T_A is invertible (Lemma 15 in T13C).

T_A is invertible **iff** T_A is both one-to-one and onto (page 8 - T13C).

T_A is both one-to-one and onto **iff** $rank(A) = n$ (Remark 6 at the end of T13C). ∎

Corollary 1

$A \in M_{n \times n}(\mathbb{F})$ is invertible **iff** $RREF(A) = I_n$.

Proof

If $RREF(A) = I_n$, then $rank(A) = n$, and thus by Lemma 2, A is invertible.

If A is invertible, then $rank(A) = n$, by Lemma 2, and $RREF(A)$ has n pivots. Thus, each of the n rows of $RREF(A)$ has a leading 1 and therefore $RREF(A) = I_n$. ■

Remark 1

Lemma 2 and Corollary 1 can be combined in the following statement:

$$A \in M_{n \times n}(\mathbb{F}) \text{is invertible } \textbf{iff } rank(A) = n \textbf{ iff } RREF(A) = I_n.$$

Lemma 3: Obtaining the inverse of a matrix (**Algorithm for Matrix Inversion**).

Suppose that $A \in M_{n \times n}(\mathbb{F})$. The following algorithm allows you to determine whether or not A is invertible, and if it is, the algorithm will provide its inverse matrix.

- Construct a super-augmented matrix $(A|I_n)$.

- Perform EROs on this matrix to reduce A in an REF.

- Once you have obtained an REF of A, then count the pivots and determine $rank(A)$.

If $rank(A) \neq n$, then A is not invertible and you can stop.

If $rank(A) = n$, then A is invertible. Obtain the RREF(A) by reducing the super-augmented matrix until it has the form $(I_n|B)$.

The second half of the super-augmented matrix, B, is equal to A^{-1}, the required inverse.

We will provide two proofs at the end of this lecture that B is equal to A^{-1} .

We continue for the time being with some examples.

Example 1

Determine if the following (2×2) matrix, A, is invertible, and if it is, then find A^{-1}.

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

Solution

We consider the super-augmented matrix $(A|I_2)$ and reduce it to REF:

$$\left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right) \xrightarrow{R_2 \to R_2 - 3R_1} \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right).$$

On the left side, we have a REF of A: since it has two pivots, then $rank(A) = 2$ and thus we know that A is invertible, so we continue reducing to RREF.

$$\xrightarrow{R_2 \to \frac{-1}{2} R_2} \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{2} & \frac{-1}{2} \end{array}\right) \xrightarrow{R_1 \to R_1 - 2R_2} \left(\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & \frac{-1}{2} \end{array}\right).$$

We thus conclude that the inverse of A is:

$$A^{-1} = \begin{pmatrix} -2 & 1 \\ \frac{3}{2} & \frac{-1}{2} \end{pmatrix} = \frac{-1}{2} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}.$$

Example 2

Determine if the following (3×3) matrix,B, is invertible, and if it is, then find B^{-1}.

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Solution

We consider the super-augmented matrix $(B|I_3)$ and reduce it to REF:

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right) \xrightarrow[R_3 \to R_3 - 7R_1]{R_2 \to R_2 - 4R_1} \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{array}\right) \xrightarrow{R_2 \to \frac{-1}{3} R_2}$$

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & \frac{-1}{3} & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{array}\right) \xrightarrow{R_3 \to R_3 + 6R_2} \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & \frac{-1}{3} & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right).$$

We can see at this point that the matrix B has rank of 2 and so it is not invertible $(n = 3)$.

Example 3

Determine if the following (3×3) matrix, C, is invertible, and if it is, then find C^{-1}.

$$C = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}.$$

Solution

We consider the super-augmented matrix $(C|I_3)$ and reduce it to REF:

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 10 & 0 & 0 & 1 \end{array}\right) \xrightarrow[R_3 \to R_3 - 7R_1]{R_2 \to R_2 - 4R_1} \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -11 & -7 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 \to \frac{-1}{3} R_2} \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & \frac{-1}{3} & 0 \\ 0 & -6 & -11 & -7 & 0 & 1 \end{array}\right) \xrightarrow{R_3 \to R3 + 6R_2} \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & \frac{-1}{3} & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right).$$

We can see at this point that the matrix C has rank of 3 and so it is invertible. We thus continue reducing to RREF.

$$\xrightarrow[R_1 \to R_1 - 3R_3]{R_2 \to R_2 - 2R_3} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & -2 & 6 & -3 \\ 0 & 1 & 0 & \frac{-2}{3} & \frac{11}{3} & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right) \xrightarrow{R_1 \to R_1 - 2R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{-2}{3} & \frac{-4}{3} & 1 \\ 0 & 1 & 0 & \frac{-2}{3} & \frac{11}{3} & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right).$$

We conclude that the inverse of C is $\begin{pmatrix} \frac{-2}{3} & \frac{-4}{3} & 1 \\ \frac{-2}{3} & \frac{11}{3} & -2 \\ 1 & -2 & 1 \end{pmatrix} = \dfrac{1}{3} \begin{pmatrix} -2 & -4 & 3 \\ -2 & 11 & -6 \\ 3 & -6 & 3 \end{pmatrix}$.

Let us check:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} \left(\frac{1}{3}\right) \begin{pmatrix} -2 & -4 & 3 \\ -2 & 11 & -6 \\ 3 & -6 & 3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} = I_3.$$

Example 4

Determine if the following (2×2) matrix, D, is invertible, and if it is, then find D^{-1}.

$$D = \begin{pmatrix} 1+i & 2-i \\ 3+2i & 4+2i \end{pmatrix}.$$

Solution

We consider the super-augmented matrix $(D|I_2)$ and reduce it to REF:

$$\left(\begin{array}{cc|cc} 1+i & 2-i & 1 & 0 \\ 3+2i & 4+2i & 0 & 1 \end{array}\right) \xrightarrow{R_1 \to \frac{1}{1+i} R_1 = \frac{1-i}{2} R_1} \left(\begin{array}{cc|cc} 1 & \frac{1-3i}{2} & \frac{1-i}{2} & 0 \\ 3+2i & 4+2i & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 \to R_2 - (3+2i) R_1} \left(\begin{array}{cc|cc} 1 & \frac{1-3i}{2} & \frac{1-i}{2} & 0 \\ 0 & \frac{-1+11i}{2} & \frac{-5+i}{2} & 1 \end{array}\right).$$

At this stage, we see that there are two pivots in the first half of this super-augmented matrix, thus D is invertible, and thus we continue.

$$\xrightarrow{R_2 \to \frac{2}{-1+11i}R_2 = -\frac{1+11i}{61}R_2} \left(\begin{array}{cc|cc} 1 & \frac{1-3i}{2} & \frac{1-i}{2} & 0 \\ 0 & 1 & \frac{8+27i}{61} & -\frac{1+11i}{61} \end{array}\right)$$

$$\xrightarrow{R_1 \to R_1 - \frac{1-3i}{2}R_2} \left(\begin{array}{cc|cc} 1 & 0 & \frac{-14-32i}{61} & \frac{17+4i}{61} \\ 0 & 1 & \frac{8+27i}{61} & -\frac{1+11i}{61} \end{array}\right).$$

We conclude that $D^{-1} = \begin{pmatrix} \frac{-14-32i}{61} & \frac{17+4i}{61} \\ \frac{8+27i}{61} & -\frac{1+11i}{61} \end{pmatrix} = \frac{1}{61}\begin{pmatrix} -14-32i & 17+4i \\ 8+27i & -1-11i) \end{pmatrix}$.

Check: $\begin{pmatrix} 1+i & 2-i \\ 3+2i & 4+2i \end{pmatrix}\left(\frac{1}{61}\right)\begin{pmatrix} -14-32i & 17+4i \\ 8+27i & -1-11i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Example 5

Consider the function $T : \mathbb{R}^2 \to \mathbb{R}^2$, given geometically by, $T_{\frac{\pi}{6}}$, a $\frac{\pi}{6}$ rotation anticlockwise, followed by, L, a reflection in the line $y = -4x$.

i) Find the standard matrix for T.
ii) Show that T is invertible.
iii) Give an expression for the inverse function of T.

Solution

We know from Topic 13C that we can obtain $[T]_S$ from the product of the two matrices: $\left[T_{\frac{\pi}{6}}\right]_S$ and $[L]_S$.

Also, we can obtain each of these matrices from Topic 13B, and so we have:

$$[T]_S = \frac{1}{17}\begin{pmatrix} -15 & -8 \\ -8 & 15 \end{pmatrix}\begin{pmatrix} \cos(\frac{\pi}{6}) & -\sin(\frac{\pi}{6}) \\ \sin(\frac{\pi}{6}) & \cos(\frac{\pi}{6}) \end{pmatrix} = \frac{1}{34}\begin{pmatrix} -8-15\sqrt{3} & 15-8\sqrt{3} \\ 15-8\sqrt{3} & 8+15\sqrt{3} \end{pmatrix}.$$

We will show that this matrix is invertible, and find its inverse, starting with $([T]_S|I_2)$:

$$\left(\begin{array}{cc|cc} \frac{-8-15\sqrt{3}}{34} & \frac{15-8\sqrt{3}}{34} & 1 & 0 \\ \frac{15-8\sqrt{3}}{34} & \frac{8+15\sqrt{3}}{34} & 0 & 1 \end{array}\right) \xrightarrow{R_1 \to \frac{34}{-8-15\sqrt{3}}R_1} \left(\begin{array}{cc|cc} 1 & \frac{15-8\sqrt{3}}{-8-15\sqrt{3}} & \frac{34}{-8-15\sqrt{3}} & 0 \\ \frac{15-8\sqrt{3}}{34} & \frac{8+15\sqrt{3}}{34} & 0 & 1 \end{array}\right)$$

$$\xrightarrow{R_2 \to R_2 - \frac{15-8\sqrt{3}}{34}R_1} \left(\begin{array}{cc|cc} 1 & \frac{15-8\sqrt{3}}{-8-15\sqrt{3}} & \frac{34}{-8-15\sqrt{3}} & 0 \\ 0 & \frac{34}{8+15\sqrt{3}} & \frac{15-8\sqrt{3}}{(8+15\sqrt{3})} & 1 \end{array}\right)$$

The $rank([T])_S = 2$, and thus $([T]_S)$ and T are invertible. We now obtain the RREF:

$$\xrightarrow{R_2 \to \frac{8+15\sqrt{3}}{34} R_2} \left(\begin{array}{cc|cc} 1 & \frac{15-8\sqrt{3}}{-8-15\sqrt{3}} & \frac{34}{-8-15\sqrt{3}} & 0 \\ 0 & 1 & \frac{15-8\sqrt{3}}{34} & \frac{8+15\sqrt{3}}{34} \end{array} \right)$$

$$\xrightarrow{R_1 \to R_1 + \frac{15-8\sqrt{3}}{8+15\sqrt{3}} R_2} \left(\begin{array}{cc|cc} 1 & 0 & \frac{-8-15\sqrt{3}}{34} & \frac{15-8\sqrt{3}}{34} \\ 0 & 1 & \frac{15-8\sqrt{3}}{34} & \frac{8+15\sqrt{3}}{34} \end{array} \right).$$

We have found $([T]_S)^{-1}$ and we can conclude that

$$T^{-1}(\mathbf{x}) = \frac{1}{34} \left(\begin{array}{cc} -8-15\sqrt{3} & 15-8\sqrt{3} \\ 15-8\sqrt{3} & 8+15\sqrt{3} \end{array} \right) \mathbf{x}.$$

Notice that $([T]_S)^{-1} = [T]_S$. There are some good reasons for this, can you think what they are?

We now prove in two ways, that if $A \in M_{n\times n}(\mathbb{F})$ with $rank(A) = n$, then, after applying the Algorithm for Matrix Inversion, the RHS of the super-augmented matrix, i.e. the matrix B is equal to A^{-1}. These proofs are both relatively straightforwards, and are examples of using material which we have already learned.

Proof I: using elementary matrices.

Let $A \in M_{n\times n}(\mathbb{F})$ be invertible, then it has rank of n and so we can row reduce it to the identity.

Suppose that we do this with EROs, $op(1), op(2), \ldots, op(q)$ and that the corresponding elementary matrices are $E_1, E_2, \ldots, E_q$.

From Lemma 6 (T12B), we have: $E_q E_{q-1} \ldots E_2 E_1 A = I_n$.

We can then deduce that $(E_q E_{q-1} \ldots E_2 E_1) = A^{-1}$ from Lemma 1. Thus

$$A^{-1} = (E_q E_{q-1} \ldots E_2 E_1) = (E_q E_{q-1} \ldots E_2 E_1) I_n = (E_q(E_{q-1} \ldots (E_2(E_1 I_n)) \ldots)).$$

Thus we can obtain A^{-1} by applying to I_n, the same EROs, $op(1), op(2), \ldots, op(q)$, that we apply to the matrix A, and in exactly the same order. $\blacksquare$

Proof II: using systems of equations.

We begin with an observation, and that is:

suppose that you have a system of equations, $C\mathbf{x} = \mathbf{k}$, where C is invertible.
If you take the augmented matrix, $(C|\mathbf{k})$ and reduce it to RREF, then you will get $(I_n|\mathbf{y})$, and $\mathbf{y}$ will be the unique solution, i.e. $C\mathbf{y} = \mathbf{k}$.

Now suppose that we have an invertible matrix $A \in M_{n\times n}(\mathbb{F})$ and $AD = I_n$.

We are trying to obtain the inverse matrix, D.
If we consider the columns of each side of the equation $AD = I_n$, we then have:

$$A\mathbf{d}_1 = \mathbf{e}_1, \quad A\mathbf{d}_2 = \mathbf{e}_2, \quad \ldots, \quad A\mathbf{d}_n = \mathbf{e}_n,$$

and so we have n systems of equations with the same coefficient matrix.

We can put these altogether into a super-augmented matrix (see T10A for another example), that is, we form:

$$(A|\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = (A|I_n).$$

We use EROs so that A is in RREF, and $(A|I_n)$ is in the form $(I_n|B)$.
Then the first column of B will be the unique solution to the first equation, $A\mathbf{d}_1 = \mathbf{e}_1$, that is, it will be $\mathbf{d}_1$.

The second column of B will be the unique solution to the second equation, $A\mathbf{d}_2 = \mathbf{e}_2$, that is, it will be $\mathbf{d}_2$, and so on.

The last column of B will be the unique solution to the last equation, $A\mathbf{d}_n = \mathbf{e}_n$, that is, it will be $\mathbf{d}_n$.

We conclude that the matrix B is equal to D, the inverse of A. $\blacksquare$

Lemma 4: The inverse of a matrix in $M_{2\times 2}(\mathbb{F})$.

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then A is invertible **iff** $ad - bc \neq 0$, and in this case,

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof

It may be shown that $rank(A) = 2$ iff $ad - bc \neq 0$. And thus A is invertible iff $ad - bc \neq 0$.
If $ad - bc \neq 0$, then we claim that the following matrix

$$\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \text{ is the inverse matrix of } A.$$

We check the above statement:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\left(\frac{1}{ad-bc}\right)\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = I_2. \qquad \blacksquare$$

Remark 2:

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $ad - bc$ is called the **determinant** of A (see the next 3 lectures).

Examples 6

Find the inverses of the following matrices if they exist.

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix} \quad D = \begin{pmatrix} 1+i & 2-i \\ 3+2i & 4+2i \end{pmatrix}$$

Solution

For A: the determinant of A is $(1(4) - 2(3)) = -2$, and so A is invertible with

$$A^{-1} = \frac{-1}{2}\begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}.$$

For B: the determinant of B is $(1(6) - 2(3)) = 0$, and so B is not invertible.

For D (see Example 5): the determinant of D is

$$((1+i)(4+2i) - (2-i)(3+2i)) = -6+5i,$$

and so D is invertible. Note that $\dfrac{1}{-6+5i} = \dfrac{-6-5i}{61}$, so that

$$D^{-1} = \frac{-6-5i}{61}\begin{pmatrix} 4+2i & -2+i \\ -3-2i & 1+i \end{pmatrix} = \frac{1}{61}\begin{pmatrix} -14-32i & 17+4i \\ 8+27i & -1-11i \end{pmatrix}.$$

Topic 15A
The Determinant I

The determinant of a square matrix is a number. We will begin by showing you how to compute this number, and then we will state some results which will enable us to simplify this computation. The main result is that a matrix is invertible iff its determinant is not zero.

Definition 1: Submatrix

Let $A \in M_{n\times n}$. We define the $(i,j)^{th}$ submatrix of A, denoted by $M_{ij}(A)$, to be the $(n-1)\times(n-1)$ matrix which one obtains from A by removing the i^{th} row from A and the j^{th} column from A.

Example 1

$$\text{Let } B = \begin{pmatrix} 4 & 6 & 8 \\ 6 & -3 & 2 \\ 5 & 7 & 9 \end{pmatrix}. \text{ Then } M_{22}(B) = \begin{pmatrix} 4 & 8 \\ 5 & 9 \end{pmatrix} \text{ and } M_{31}(B) = \begin{pmatrix} 6 & 8 \\ -3 & 2 \end{pmatrix}.$$

Definition 2: Determinant of a 1×1 and 2×2 matrix.

If $A \in M_{1\times 1}(\mathbb{F})$, then the determinant of A, written $\det(A)$, is: $\det(A) = a_{11}$.
If $A \in M_{2\times 2}(\mathbb{F})$, then the determinant of A, written $\det(A)$, is: $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

Example 2

Let $B = (1-2i)$. Then, $\det(B) = 1-2i$.

Let $C = \begin{pmatrix} 6 & 8 \\ -3 & 2 \end{pmatrix}$. Then, $\det(C) = 6(2) - 8(-3) = 36$.

Definition 3: First row expansion of the determinant.

Let $A \in M_{n\times n}(\mathbb{F})$ with $n \geq 2$.

We define the **determinant** function, $\det : M_{n\times n} \to \mathbb{F}$, by:

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

This expression is known as the first row expansion of the determinant.
The expression requires you to multiply entries on the **first row** of the matrix A by ± 1, and then by the determinant of a submatrix.

For a 3×3 matrix, this will involve the evaluation of 3 determinants of 2×2 matrices, which we can already do.
For a 4×4 matrix this will involve the evaluation of 4 determinants of 3×3 matrices, each of which involve the evaluation of 3 determinants of 2×2 matrices.

Do not despair, we will learn some techniques in Topic 15B, which will permit us to avoid too many tedious computations, if we wish.

Examples 3

What is the determinant of the matrix $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$?

Solution

$\det(B) = \sum_{j=1}^{j=n} b_{1j}(-1)^{1+j} \det(M_{1j}(B)), \quad$ that is :

$$\det(B) = b_{11}(-1)^{1+1}\det(M_{11}(B)) + b_{12}(-1)^{1+2}\det(M_{12}(B)) + b_{13}(-1)^{1+3}\det(M_{13}(B))$$

$$= (1)(1)\det\left(\begin{pmatrix} 5 & 6 \\ 8 & 10 \end{pmatrix}\right) + (2)(-1)\det\left(\begin{pmatrix} 4 & 6 \\ 7 & 10 \end{pmatrix}\right) + (3)(1)\det\left(\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}\right)$$

$$= 1(50 - 48) - 2(40 - 42) + 3(32 - 35) = -3.$$

What is the determinant of the matrix $C = \begin{pmatrix} 1 & 2i & 3 \\ 4-i & 5+2i & 6 \\ 7+3i & 8-2i & 10 \end{pmatrix}$?

Solution

$\det(C) = \sum_{j=1}^{j=n} c_{1j}(-1)^{1+j} \det(M_{1j}(C))$, that is :

$$\det(C) = c_{11}(-1)^{1+1}\det(M_{11}(C)) + c_{12}(-1)^{1+2}\det(M_{12}(C)) + c_{13}(-1)^{1+3}\det(M_{13}(C))$$

$$= (1)(1)\det\left(\begin{pmatrix} 5+2i & 6 \\ 8-2i & 10 \end{pmatrix}\right) + (2i)(-1)\det\left(\begin{pmatrix} 4-i & 6 \\ 7+3i & 10 \end{pmatrix}\right) +$$
$$(3)(1)\det\left(\begin{pmatrix} 4-i & 5+2i \\ 7+3i & 8-2i \end{pmatrix}\right)$$
$$= 1(50 + 20i - (48 - 12i)) - 2i(40 - 10i - (42 + 18i)) + 3(30 - 16i - (29 + 29i))$$

$$= -51 - 99i.$$

One of the remarkable features of the determinant is that we may evaluate it by by moving along any row, not just the first one.

Definition 4: I^{th} row expansion of the determinant.

Let $A \in M_{n\times n}(\mathbb{F})$ with $n \geq 2$, the I^{th} row expansion of the determinant is obtained by performing the following calculations:

$$\det(A) = \sum_{j=1}^{j=n} a_{Ij}(-1)^{I+j}\det(M_{Ij}(A)).$$

The expression requires you to multiply entries on the I^{th} row of the matrix A by ± 1, and then by the determinant of a submatrix.

Examples 4

Evaluate the determinant of the matrix $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$ using the second row expansion.

Solution

$\det(B) = \sum\limits_{j=1}^{j=n} b_{2j}(-1)^{2+j}\det(M_{2j}(B))$, that is :

$$\det(B) = b_{21}(-1)^{2+1}\det(M_{21}(B)) + b_{22}(-1)^{2+2}\det(M_{22}(B)) + b_{23}(-1)^{2+3}\det(M_{23}(B))$$

$$= (4)(-1)\det\left(\begin{pmatrix} 2 & 3 \\ 8 & 10 \end{pmatrix}\right) + (5)(1)\det\left(\begin{pmatrix} 1 & 3 \\ 7 & 10 \end{pmatrix}\right) + (6)(-1)\det\left(\begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}\right)$$

$$= -4(20-24) + 5(10-21) - 6(8-14) = -3.$$

Evaluate the determinant of the matrix $C = \begin{pmatrix} 1 & 2i & 3 \\ 4-i & 5+2i & 6 \\ 7+3i & 8-2i & 10 \end{pmatrix}$ using a third row expansion.

Solution

$\det(C) = \sum\limits_{j=1}^{j=n} c_{3j}(-1)^{3+j}\det(M_{3j}(C))$, that is :

$$\det(C) = c_{31}(-1)^{3+1}\det(M_{31}(C)) + c_{32}(-1)^{3+2}\det(M_{32}(C)) + c_{33}(-1)^{3+3}\det(M_{33}(C))$$

$$= (7+3i)(1)\det\left(\begin{pmatrix} 2i & 3 \\ 5+2i & 6 \end{pmatrix}\right) + (8-2i)(-1)\det\left(\begin{pmatrix} 1 & 3 \\ 4-i & 6 \end{pmatrix}\right) +$$

$$(10)(1)\det\left(\begin{pmatrix} 1 & 2i \\ 4-i & 5+2i \end{pmatrix}\right)$$

$$= (7+3i)(-15+6i) + (-8+2i)(-6+2i) + 10(3-6i) = -51-99i.$$

A further remarkable features of the determinant is that it may be evaluated by moving down any column.

Definition 5: J^{th} column expansion of the determinant.

Let $A \in M_{n\times n}(\mathbb{F})$ with $n \geq 2$, the J^{th} column expansion of the determinant is obtained by performing the following calculation:

$$\det(A) = \sum_{i=1}^{i=n} a_{iJ}(-1)^{i+J}\det(M_{_{iJ}}(A)).$$

The expression requires you to multiply entries on the J^{th} column of the matrix A by ± 1, and then by the determinant of a submatrix.

Examples 5

Evaluate the determinant of the matrix $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$ using a third column expansion.

Solution

$\det(B) = \sum\limits_{i=1}^{i=n} b_{i3}(-1)^{i+3}\det(M_{i3}(B))$, that is :

$$\det(B) = b_{13}(-1)^{1+3}\det(M_{13}(B)) + b_{23}(-1)^{2+3}\det(M_{23}(B)) + b_{33}(-1)^{3+3}\det(M_{33}(B))$$

$$= (3)(1)\det\left(\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}\right) + (6)(-1)\det\left(\begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}\right) + (10)(1)\det\left(\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}\right)$$

$$= 3(32-35) - 6(8-14) + 10(5-8) = -3.$$

Evaluate the determinant of the matrix $C = \begin{pmatrix} 1 & 2i & 3 \\ 4-i & 5+2i & 6 \\ 7+3i & 8-2i & 10 \end{pmatrix}$ using the second column expansion.

Solution

$\det(C) = \sum_{i=1}^{i=n} c_{i2}(-1)^{i+2}\det(M_{i2}(C))$, that is :

$$\det(C) = c_{12}(-1)^{1+2}\det(M_{12}(C)) + c_{22}(-1)^{2+2}\det(M_{22}(C)) + c_{32}(-1)^{3+2}\det(M_{32}(C))$$

$$= (2i)(-1)\det\left(\begin{pmatrix} 4-i & 6 \\ 7+3i & 10 \end{pmatrix}\right) + (5+2i)(1)\det\left(\begin{pmatrix} 1 & 3 \\ 7+3i & 10 \end{pmatrix}\right) +$$

$$(8-2i)(-1)\det\left(\begin{pmatrix} 1 & 3 \\ 4-i & 6 \end{pmatrix}\right)$$

$$= (-2i)(-2-28i)) + (5+2i)(-11-9i) + (-8+2i)(-6+3i) = -51-99i.$$

Thus you may choose the row or column that you want to do the expansion along or down. In practice, it is usually easier to choose a row or column with many zeros, since then you do not need to evaluate the corresponding determinants of smaller matrices.

Example 6

Evaluate the determinant of $W = \begin{pmatrix} 1 & 3 & 0 & 5 \\ 0 & 2 & 0 & 7 \\ 567 & 234 & 14 & 235 \\ 4 & 8 & 0 & 3 \end{pmatrix}$.

Solution

We choose to perform a third column expansion.

$$\det(W) = \sum_{i=1}^{i=4} w_{i3}(-1)^{i+3}\det(M_{i3}(W))$$

$$\det(W) = 0(1)\det(M_{13}(W)) + 0(-1)\det(M_{23}(W)) + 14(1)\det(M_{33}(W)) + 0(-1)\det(M_{43}(W))$$

$$= 14\det\left(\begin{pmatrix} 1 & 3 & 5 \\ 0 & 2 & 7 \\ 4 & 8 & 3 \end{pmatrix}\right)$$

$$= 14\left[(1)\det\left(\begin{pmatrix} 2 & 7 \\ 8 & 3 \end{pmatrix}\right) + (4)\det\left(\begin{pmatrix} 3 & 5 \\ 2 & 7 \end{pmatrix}\right)\right] = -84.$$

Remark 1

Note that if a square matrix has either a row or column that consists only of zeros, then that matrix has a determinant of zero.

Definition 6: Cofactor

Let $A \in M_{n\times n}(\mathbb{F})$. We define the $(i,j)^{th}$ **cofactor** of A, denoted by $C_{ij}(A)$: that is,

$$C_{ij}(A) = (-1)^{i+j}\det(M_{ij}(A)).$$

This is a useful piece of notation as it simplifies the expression for the determinant. It now follows that we can write the determinant of A as either:

$$\det(A) = \sum_{j=1}^{j=n} a_{Ij}C_{Ij}(A),\ \ (I^{th} \text{ row expansion})$$

or

$$\det(A) = \sum_{i=1}^{i=n} a_{iJ}C_{iJ}(A)\ \ (J^{th} \text{ column expansion})$$

Lemma 1

Let $A \in M_{n\times n}(\mathbb{F})$. Then, $\det(A^T) = \det(A)$.

Proof

The two main features of the proof are that:

$$(A^T)_{ij} = (A)_{ji} = a_{ji} \quad (*) \quad \text{and} \quad M_{ij}(A^T) = (M_{ji}(A))^T \quad (**).$$

Let $P(n)$ be the statement:

$$\text{If } A_n \in M_{n\times n}, \text{ then } \det((A_n)^T) = \det(A_n), \text{ for all integers } n \geq 1.$$

$P(1)$ states that if $A_1 \in M_{1\times 1}$, then $\det((A_1)^T) = \det(A_1)$.
This is true, since they are both equal to a_{11}.

Now suppose that $P(k)$ is true for some integer $k \geq 1$, so the inductive hypothesis is:

$$\text{If } A_k \in M_{k\times k}\ , \text{ then } \det((A_k)^T) = \det(A_k),\ \text{ for some integer } k \geq 1.$$

We now examine the statement $P(k+1)$ which states that:

$$\text{If } A_{(k+1)} \in M_{(k+1)\times(k+1)}\ , \text{ then } \det((A_{(k+1)})^T) = \det(A_{(k+1)}).$$

If we perform a **first row** expansion to find the determinant of $(A_{(k+1)})^T$, we have:

$$\begin{aligned}\det((A_{(k+1)})^T) &= \sum_{j=1}^{j=k+1} (A_{(k+1)})^T)_{1j} \ (-1)^{1+j} \ \det\left(M_{1j}((A_{(k+1)})^T)\right) \\ &= \sum_{j=1}^{j=k+1} (A_{(k+1)})_{j1} \ (-1)^{j+1} \ \det\left((M_{j1}(A_{(k+1)}))^T\right), \quad \text{by}(*) \text{ and } (**).\end{aligned}$$

Since $M_{j1}(A_{(k+1)}) \in M_{k\times k,}$ we can use the inductive hypothesis to write:

$$\det(M_{j1}(A_{(k+1)}))^T = \det\left(M_{j1}(A_{(k+1)})\right),$$

and so we have:

$$\det((A_{(k+1)})^T) = \sum_{j=1}^{j=k+1} (A_{(k+1)})_{j1} \ (-1)^{j+1} \ \det\left(M_{j1}(A_{(k+1)})\right),$$

which is the **first column** expansion of the determinant of the matrix $A_{(k+1)}$, and thus

$$\det((A_{(k+1)})^T) = \det(A_{(k+1)}).$$

We conclude that $P(k+1)$ is a true statement and thus by the Principle of Mathematical Induction, that $P(n)$ is true for all integers $n \geq 1$. ■

Lemma 2

Let $A \in M_{n\times n}(\mathbb{F})$. and suppose that A is upper (lower) triangular, then

$$\det(A) = a_{11}a_{22}\dots a_{nn} = \prod_{i=1}^{n} a_{ii}.$$

Proof by induction

Let $P(n)$ be the statement :

If $A_n \in M_{n\times n}$ is upper triangular, then $\det(A_n) = a_{11}a_{22}\dots a_{nn}$, for all integers $n \geq 1$.

$P(1)$ states that if $A_1 \in M_{1\times 1}$ is upper triangular, then $\det(A_1) = a_{11}$, which is true.

Now suppose that $P(k)$ is true for some integer $k \geq 1$, that is:

If $A_k \in M_{k\times k}$ is upper triangular, then $\det(A_k) = a_{11}a_{22}\dots a_{kk}$.

We now examine the statement $P(k+1)$ which states that :

If $A_{(k+1)} \in M_{(k+1)\times(k+1)}$ is upper triangular, then $\det(A_{(k+1)}) = a_{11}a_{22}\ldots a_{(k+1)(k+1)}$.

Suppose we perform a $(k+1)^{th}$ row expansion of this determinant.

All the entries in the last row of $A_{(k+1)}$ are zero except the $((k+1),(k+1))^{th}$, which is $a_{(k+1)(k+1)}$.

We thus have that:

$$\det(A_{(k+1)}) = a_{(k+1)(k+1)}(-1)^{(k+1)+(k+1)}\det(M_{_{(k+1)(k+1)}}(A_{(k+1)})).$$

However, $M_{_{(k+1)(k+1)}}(A_{(k+1)})$ is a $k \times k$ upper triangular matrix, the matrix obtained from $A_{(k+1)}$ by removing its last row and last column.

Thus, by the inductive hypothesis $P(k)$, we get:

$$\det(M_{_{(k+1)(k+1)}}(A_{(k+1)})) = a_{11}a_{22}\ldots a_{kk}.$$

We then have:

$$\det(A_{(k+1)}) = a_{(k+1)(k+1)}a_{11}a_{22}\ldots a_{kk} = a_{11}a_{22}\ldots a_{(k+1)(k+1)}.$$

We conclude that $P(k+1)$ is a true statement and thus by the Principle of Mathematical Induction, that $P(n)$ is true for all integers $n \geq 1$. ■

Corollary 1 (of Lemma 2)

Let $A \in M_{n\times n}(\mathbb{F})$. If A is diagonal, then $\det(A) = a_{11}a_{22}\ldots a_{nn} = \prod_{i=1}^{n} a_{ii}$.

In particular, $\det(I_n) = 1$.

The proof is immediate from Lemma 2.

We conclude this lecture with a theorem, and a few corollaries, the consequences of which will provide the foundations for the material for the next topic.

Theorem 1

Let $A \in M_{n\times n}(\mathbb{F})$ and suppose that we write $A = \begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^i \\ \vdots \\ \mathbf{A}^k \\ \vdots \\ \mathbf{A}^n \end{pmatrix}$, then

a) The determinant is skew-symmetric under the interchange of rows, that is:

$$\det\left(\begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^k \\ \vdots \\ \mathbf{A}^i \\ \vdots \\ \mathbf{A}^n \end{pmatrix}\right) = -\det\left(\begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^i \\ \vdots \\ \mathbf{A}^k \\ \vdots \\ \mathbf{A}^n \end{pmatrix}\right)$$

b) The determinant is a linear operation on rows, by which we mean that if $\mathbf{B}^i \in M_{1\times n}$ and if c_1, c_2 are scalars, then:

$$\det\left(\begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ c_1\mathbf{A}^i + c_2\,\mathbf{B}^i \\ \vdots \\ \mathbf{A}^k \\ \vdots \\ \mathbf{A}^n \end{pmatrix}\right) = c_1 \det\left(\begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^i \\ \vdots \\ \mathbf{A}^k \\ \vdots \\ \mathbf{A}^n \end{pmatrix}\right) + c_2 \det\left(\begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{B}^i \\ \vdots \\ \mathbf{A}^k \\ \vdots \\ \mathbf{A}^n \end{pmatrix}\right).$$

We do not prove these results in this course.

Remark 2

The same statement can be made if the word column replaces the word row throughout.

Corollary 2

Let $A \in M_{n\times n}(\mathbb{F})$. If A has two identical rows(columns), then $\det(A) = 0$.

Proof

If you switch the two identical rows(columns) and use Theorem 1,
then you have $\det(A) = -\det(A)$, and so we conclude that $\det(A) = 0$. $\blacksquare$

Corollary 3: Determinants of elementary matrices.

Let E_i be an elementary matrix of type k then

(i) $\det(E_1) = -1$, where E_1 is obtained from I_n by interchanging 2 rows. $(k = \text{I})$

(ii) $\det(E_2) = m$, where E_2 is obtained from I_n by scaling a row by $m \neq 0$. $(k = \text{II})$

(iii) $\det(E_3) = 1$, where E_3 is obtained from I_n by adding a non-zero multiple of a row to another row $(k = \text{III})$.

Proof

We already know that $\det(I_n) = 1$, and so

(i) $\det(E_1) = -1$ follows from part a) of Theorem 1.

(ii) Follows from part b) of Theorem 1, let $c_1 = m$ and $c_2 = 0$

(iii) Follows from part b) of Theorem 1 and Corollary 2, let $c_1 = 1$ and $\mathbf{B}^i = \mathbf{A}^k$. $\blacksquare$

Topic 15B

The Determinant II

Corollary 4: EROs and the determinant.

Let $A \in M_{n\times n}(\mathbb{F})$ and suppose we perform a single ERO on A to produce the matrix B.

(i) If the ERO is type I ($R_i \leftrightarrow R_j$), then $\det(B) = -\det(A)$.

(ii) If the ERO is type II ($R_i \to mR_i, m \neq 0$), then $\det(B) = m\det(A)$.

(iii) If the ERO is type III ($R_i \to R_i + mR_j, i \neq j$), then $\det(B) = \det(A)$.

Proof

Part (i) is part a) of Theorem 1 (T15A).
Part (ii) is part b) of Theorem 1 (T15A) with $c_1 = m$ and $c_2 = 0$.
Part (iii) is part b) of Theorem 1 (T15A) with $\mathbf{B}^i = \mathbf{A}^k$, $c_1 = 1$, $c_2 = m$, and using Corollary 2 (T15A). ■

Example 7

Consider the following matrix B. In Example 5 (T15A), it was shown that $\det(B) = -3$.

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$$

We will verify Corollary 4 when we apply each of the three types of EROs on B.
The determinants in Example 7 are always evaluated by expanding along the first row.

The matrix C_1 is obtained from B by performing the ERO: $R_2 \leftrightarrow R_3$.

$$C_1 = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 10 \\ 4 & 5 & 6 \end{pmatrix}.$$

$$\begin{aligned}\det(C_1) &= 1(1)\det\left(\begin{pmatrix} 8 & 10 \\ 5 & 6 \end{pmatrix}\right) + 2(-1)\det\left(\begin{pmatrix} 7 & 10 \\ 4 & 6 \end{pmatrix}\right) + 3(1)\det\left(\begin{pmatrix} 7 & 8 \\ 4 & 5 \end{pmatrix}\right) \\ &= (-2) - 4 + 9 = 3.\end{aligned}$$

That is, $\det(C_1) = -\det(B)$.

The matrix C_2 is obtained from B by performing the ERO: $R_2 \to (-4)R_2$.

$$C_2 = \begin{pmatrix} 1 & 2 & 3 \\ (-4)(4) & (-4)(5) & (-4)(6) \\ 7 & 8 & 10 \end{pmatrix}.$$

$$\det(C_2) = 1(1)\det\left(\begin{pmatrix} (-4)(5) & (-4)(6) \\ 8 & 10 \end{pmatrix}\right) + 2(-1)\det\left(\begin{pmatrix} (-4)(4) & (-4)(6) \\ 7 & 10 \end{pmatrix}\right) +$$
$$3(1)\det\left(\begin{pmatrix} (-4)(4) & (-4)(5) \\ 7 & 8 \end{pmatrix}\right)$$
$$= (-200+192) - 2(-160+168) + 3(-128+140) = -8 - 2(8) + 3(12) = 12.$$

That is, $\det(C_2) = (-4)\det(B)$.

The matrix C_3 is obtained from B by performing the ERO: $R_1 \to R_1 + 6R_3$.

$$C_3 = \begin{pmatrix} 43 & 50 & 63 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}.$$

$$\det(C_3) = 43(1)\det\left(\begin{pmatrix} 5 & 6 \\ 8 & 10 \end{pmatrix}\right) + 50(-1)\det\left(\begin{pmatrix} 4 & 6 \\ 7 & 10 \end{pmatrix}\right) + 63(1)\det\left(\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}\right)$$
$$= 43(2) - 50(-2) + 63(-3) = -3.$$

That is $\det(C_3) = \det(B)$.

Corollary 5

Let $A \in M_{n\times n}(\mathbb{F})$ and suppose we perform a single ERO on A to produce the matrix B. Assume that the corresponding elementary matrix is E. Then

$$\det(B) = \det(E)\det(A).$$

The proof is obtained by combining Corollaries 3 (T15A) and 4, and considering the three types of elementary row operations one by one.

Corollary 6

Let $A \in M_{n\times n}(\mathbb{F})$ and suppose we perform a sequence of EROs $op(1), op(2), \ldots, op(q)$ on the matrix A to obtain the matrix B.

Suppose that the elementary matrix corresponding to ERO k is labelled E_k. We then have:

$$\det(B) = \det(E_q E_{q-1} \dots E_1 A) = \det(E_q)\det(E_{q-1})\dots\det(E_2)\det(E_1)\det(A).$$

The proof is by induction and Corollary 5.

How we evaluate determinants in practice.

We can make use of the Corollary 6 to compute the determinant of a matrix with relative ease. The idea is to perform elementary row operations on your matrix, A say, until you obtain another matrix, B say, whose determinant is easily obtained, for example, if you perform Gaussian elimination, then B could have a lot of zeros.

You have to keep track of which elementary row operations that you do, so that you can relate the determinants of A and B. The good news is that type III operations, which we tend to do the most of, do not affect the determinant. Any type I operation performed will introduce a negative sign. We rarely perform type II operations. However, Theorem 1 part b) (T15A) allows us to factor out a scalar from a row or a column.

Example 8

Evaluate the determinant of $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$.

Solution

$\det\left(\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}\right)$, perform $\begin{matrix} R_2 \to R_2 - 4R_1 \\ R_3 \to R_3 - 7R_1 \end{matrix}$ which do not change the det

$= \det\left(\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{pmatrix}\right)$, perform $R_3 \to R_3 - 2R_2$ which does not change the det

$= \det\left(\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{pmatrix}\right) = (1)(-3)(1) = -3$ (the last matrix is upper triangular).

Evaluate the determinant of $C = \begin{pmatrix} 1 & 2i & 3 \\ 4-i & 5+2i & 6 \\ 7+3i & 8-2i & 10 \end{pmatrix}$.

Solution

$\det\left(\begin{pmatrix} 1 & 2i & 3 \\ 4-i & 5+2i & 6 \\ 7+3i & 8-2i & 10 \end{pmatrix}\right)$, perform $R_2 \to R_2 - 2R_1$, (no change in det)

$$= \det\left(\begin{pmatrix} 1 & 2i & 3 \\ 2-i & 5-2i & 0 \\ 7+3i & 8-2i & 10 \end{pmatrix}\right), \text{ perform } R_3 \to R_3 - \tfrac{10}{3}R_1, \text{ (no change in det)}$$

$$= \det\left(\begin{pmatrix} 1 & 2i & 3 \\ 2-i & 5-2i & 0 \\ \frac{11}{3}+3i & 8-\frac{26}{3}i & 0 \end{pmatrix}\right), \text{ expand the det along the third column}$$

$$= 3\det\left(\begin{pmatrix} 2-i & 5-2i \\ \frac{11}{3}+3i & 8-\frac{26}{3}i \end{pmatrix}\right) = 3\left(\tfrac{22}{3} - \tfrac{76}{3}i - \left(\tfrac{73}{3} - \tfrac{23}{3}i\right)\right) = -51-99i$$

The next example looks pretty nasty at the start, but the final answer pops out after 15 elementary row operations, instead of evaluating 5 determinants of 4×4 matrices.

Example 9

Find the determinant of the matrix $F = \begin{pmatrix} 2 & 3 & -6 & 5 & 9 \\ 4 & 1 & 4 & -7 & 6 \\ 7 & -4 & 4 & 3 & 2 \\ 9 & 19 & 16 & 2 & 13 \\ 6 & 4 & -6 & 18 & 3 \end{pmatrix}$.

Solution

$$\det\left(\begin{pmatrix} 2 & 3 & -6 & 5 & 9 \\ 4 & 1 & 4 & -7 & 6 \\ 7 & -4 & 4 & 3 & 2 \\ 9 & 19 & 16 & 2 & 13 \\ 6 & 4 & -6 & 18 & 3 \end{pmatrix}\right), \text{ factor (2) from } R_1$$

$$= (2)\det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 4 & 1 & 4 & -7 & 6 \\ 7 & -4 & 4 & 3 & 2 \\ 9 & 19 & 16 & 2 & 13 \\ 6 & 4 & -6 & 18 & 3 \end{pmatrix}\right), \quad \begin{matrix} R_2 \to R_2 - 4R_1 \\ R_3 \to R_3 - 7R_1 \\ R_4 \to R_4 - 9R_1 \\ R_5 \to R_5 - 6R_1 \end{matrix}$$

$$= 2\det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 0 & -5 & 16 & -17 & -12 \\ 0 & \frac{-29}{2} & 25 & \frac{-29}{2} & \frac{-59}{2} \\ 0 & \frac{11}{2} & 43 & \frac{-41}{2} & \frac{-55}{2} \\ 0 & -5 & 12 & 3 & -24 \end{pmatrix}\right), \quad \begin{matrix} \text{factor } (-5) \text{ from } R_2 \\ \text{factor } (\frac{1}{2}) \text{ from } R_3 \\ \text{factor } (\frac{1}{2}) \text{ from } R_4 \end{matrix}$$

$$= 2(-5)(\tfrac{1}{2})(\tfrac{1}{2})\det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 0 & 1 & \frac{-16}{5} & \frac{17}{5} & \frac{12}{5} \\ 0 & -29 & 50 & -29 & -59 \\ 0 & 11 & 86 & -41 & -55 \\ 0 & -5 & 12 & 3 & -24 \end{pmatrix}\right), \quad \begin{matrix} R_3 \to R_3 + 29R_2 \\ R_4 \to R_4 - 11R_2 \\ R_5 \to R_5 + 5R_2 \end{matrix}$$

$$= \left(\tfrac{-5}{2}\right) \det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 0 & 1 & \frac{-16}{5} & \frac{17}{5} & \frac{12}{5} \\ 0 & 0 & \frac{-214}{5} & \frac{348}{5} & \frac{53}{5} \\ 0 & 0 & \frac{606}{5} & \frac{-392}{5} & \frac{-407}{5} \\ 0 & 0 & -4 & 20 & 12 \end{pmatrix}\right), \text{ factor } \left(\tfrac{-214}{5}\right) \text{ from } R_3$$

$$= \left(\tfrac{-5}{2}\right)\left(\tfrac{-214}{5}\right) \det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 0 & 1 & \frac{-16}{5} & \frac{17}{5} & \frac{12}{5} \\ 0 & 0 & 1 & \frac{-348}{214} & \frac{-53}{214} \\ 0 & 0 & \frac{606}{5} & \frac{-392}{5} & \frac{-407}{5} \\ 0 & 0 & -4 & 20 & 12 \end{pmatrix}\right), \quad \begin{matrix} R_4 \to R_4 - \frac{606}{5}R_3 \\ R_5 \to R_5 + 4R_3 \end{matrix}$$

$$= (107) \det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 0 & 1 & \frac{-16}{5} & \frac{17}{5} & \frac{12}{5} \\ 0 & 0 & 1 & \frac{-348}{214} & \frac{-53}{214} \\ 0 & 0 & 0 & \frac{12700}{107} & \frac{-5498}{107} \\ 0 & 0 & 0 & \frac{1444}{107} & \frac{-1390}{107} \end{pmatrix}\right), \text{ factor } \left(\tfrac{12700}{107}\right) \text{ from } R_4$$

$$= (107)\left(\tfrac{12700}{107}\right) \det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 0 & 1 & \frac{-16}{5} & \frac{17}{5} & \frac{12}{5} \\ 0 & 0 & 1 & \frac{-348}{214} & \frac{-53}{214} \\ 0 & 0 & 0 & 1 & \frac{-5498}{12700} \\ 0 & 0 & 0 & \frac{1444}{107} & \frac{-1390}{107} \end{pmatrix}\right), R_5 \to R_5 - \tfrac{1444}{107}R_4$$

$$= (12700) \det\left(\begin{pmatrix} 1 & \frac{3}{2} & -3 & \frac{5}{2} & \frac{9}{2} \\ 0 & 1 & \frac{-16}{5} & \frac{17}{5} & \frac{12}{5} \\ 0 & 0 & 1 & \frac{-348}{214} & \frac{-53}{214} \\ 0 & 0 & 0 & 1 & \frac{-5498}{12700} \\ 0 & 0 & 0 & 0 & \frac{-22696}{3175} \end{pmatrix}\right) = (12700)\left(\tfrac{-22696}{3175}\right) = -90784 = \det(F).$$

In the next few results we make us of the following idea. Suppose we perform a sequence of EROs, $op(1), op(2), \ldots, op(q)$ on an $n \times n$ matrix A in order to obtain $RREF(A)$, which we refer to as R, so $R = RREF(A)$.

Suppose that the elementary matrix corresponding to ERO k is labelled E_k, then we have:

$$R = E_q E_{q-1} \ldots E_2 E_1 A.$$

Also, each elementary matrix is invertible with $E_i^{-1} = F_i$, for $i = 1, \ldots, q$, which is also another elementary matrix. Thus, we can solve for A and get:

$$A = F_1 F_2 \dots F_{q-1} F_q R.$$

Note that there are two possibilities for R, and these are either, using Corollary 1 (T14):

(i) $R = I_n$, and so A is invertible

or

(ii) $R \neq I_n$, and so A is not invertible, in this case R has a row of zeros.

Corollary 7: Invertibility iff the determinant is non-zero.

Let $A \in M_{n\times n}(\mathbb{F})$.

A is invertible **iff** $\det(A) \neq 0$.

Proof

Corollary 6 gives us that:

$$\det(A) = \det(F_1)\det(F_2)\dots\det(F_{q-1})\det(F_q)\det(R).$$

We know that $\det(F_i) \neq 0$, for $i = 1, 2, \dots, q$.

There are two possibilities and using Corollary 1 (T14), we conclude:

(i) A is invertible, and so $R = I_n$. Clearly then, $\det(A) \neq 0$.

or

(ii) A is not invertible, and so $R \neq I_n$ and must have a row of zeros.

In this case $\det(R) = 0$, and we conclude that $\det(A) = 0$ too. ■

Example 10

For what value of the constant k is the matrix $B = \begin{pmatrix} 3 & 5 & 7 \\ 6 & k & 14 \\ 2 & 4 & 6 \end{pmatrix}$ singular?

Solution

Let us evaluate the determinant of B :

$$\det\left(\begin{pmatrix} 3 & 5 & 7 \\ 6 & k & 14 \\ 2 & 4 & 6 \end{pmatrix}\right) = \det\left(\begin{pmatrix} 3 & 5 & 7 \\ 0 & k-10 & 0 \\ 2 & 4 & 6 \end{pmatrix}\right), \quad R_2 - 2R_1$$

$$= 2\det\left(\begin{pmatrix} 3 & 5 & 7 \\ 0 & k-10 & 0 \\ 1 & 2 & 3 \end{pmatrix}\right), \quad \text{factor (2) from } R_3$$

$$= 2\det\left(\begin{pmatrix} 0 & -1 & -2 \\ 0 & k-10 & 0 \\ 1 & 2 & 3 \end{pmatrix}\right), \quad R_1 - 3R_3$$

$$= 4(k-10), \quad \text{expanding the determinant along the third row.}$$

Since $\det(B) = 0$ iff $k = 10$, we conclude that the matrix B is singular iff $k = 10$.

Corollary 8: Determinant of a product.

Let $A, B \in M_{n\times n}(\mathbb{F})$, then $\det(AB) = \det(A)\det(B)$.

Proof

We make use of that fact that we can write the matrix A as:

$A = F_1F_2\ldots F_{q-1}F_qR$ so that $AB = F_1F_2.\ldots F_{q-1}F_qRB$.

There are two possibilities and using Corollary 1 (T14), we conclude:

(i) If A is invertible, then $R = I_n$ and we have:

$$\begin{aligned} \det(AB) &= \det(F_1F_2\ldots F_{q-1}F_qIB) \\ &= \det(F_1)\det(F_2)\ldots\det(F_q)\det(I)\det(B), \quad \text{using Corollary 6} \\ &= \det(F_1)\det(F_2)\ldots\det(F_q)\det(B) \\ &= \det(F_1F_2\ldots F_{q-1}F_q)\det(B), \quad \text{using Corollary 6 in reverse} \\ &= \det(F_1F_2\ldots F_{q-1}F_qI)\det(B) \\ &= \det(A)\det(B). \end{aligned}$$

(ii) If A is not invertible, then $R \neq I_n$ and then R has a row of zeros. In this case, RB also has a row of zeros and so $\det(RB) = 0$.

$$\begin{aligned} \det(AB) &= \det(F_1F_2\ldots F_{q-1}F_qRB) \\ &= \det(F_1)\det(F_2)\ldots\det(F_q)\det(RB) = 0, \quad \text{using Corollary 6.} \end{aligned}$$

We also know that $\det(A) = 0,$ from Corollary 7.

So we conclude that $\det(AB) = \det(A)\det(B) = 0$.

And thus in either cases, $\det(AB) = \det(A)\det(B)$. ■

Example 11

If $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 2i & 3 \\ 4-i & 5+2i & 6 \\ 7+3i & 8-2i & 10 \end{pmatrix}$, then what is $\det(CB)$?

Solution

We already know from Topic 15A that $\det(B) = -3$ and that $\det(C) = -51 - 99i$.

We conclude that $\det(CB) = \det(C)\det(B) = (-51 - 99i)(-3) = 9(17 + 33i) = 153 + 297i$.

Example 12

Let $A, B \in M_{n\times n}(\mathbb{F})$. Prove that AB is singular iff BA is singular.

Proof

$$\begin{aligned} AB \text{ is singular} &\text{ iff } \det(AB) = 0, \quad \text{by Corollary 7} \\ &\text{ iff } \det(A)\det(B) = 0, \quad \text{by Corollary 8} \\ &\text{ iff } \det(B)\det(A) = 0 \\ &\text{ iff } \det(BA) = 0, \quad \text{by Corollary 8} \\ &\text{ iff } BA \text{ is singular}, \quad \text{by Corollary 7.} \qquad \blacksquare \end{aligned}$$

Corollary 9

Let $A \in M_{n\times n}(\mathbb{F})$ be invertible. Then $\det(A^{-1}) = (\det(A))^{-1}$.

Proof

We have that $A^{-1}A = I_n$. Thus taking the determinant of both sides, we get:

$\det(A^{-1}A) = \det(I_n)$, so that $\det(A^{-1})\det(A) = 1$, using Corollary 8.

Since A is invertible, we know that $\det(A) \neq 0$, and we can thus write:

$\det(A^{-1}) = (\det(A))^{-1}$. $\blacksquare$

Example 13

Prove that $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$ is invertible and find $\det(B^{-1})$.

Solution

We have already shown that $\det(B) = -3$, and so we know from Corollary 7 that B is invertible.

Then using Corollary 9, we get that $\det(B^{-1}) = \frac{-1}{3}$.

Topic 15C
The Determinant III - Applications

Definition 7: Adjoint (adjunct) of a matrix.

Let $A \in M_{n\times n}(\mathbb{F})$. We define the **adjoint** or **adjunct** of A, denoted $\mathrm{adj}(A)$, by:

$$(\mathrm{adj}(A))_{ij} = C_{ji}(A), \quad \text{for all } i,\ j = 1, \ldots, n.$$

That is, the adjoint (adjunct) of A is the **transpose** of the matrix of all the cofactors of A.

Example 14

Consider the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$.

Note we write M_{ij} and C_{ij} instead of $M_{ij}(A)$ and $C_{ij}(A)$.

$$\begin{array}{llll}
M_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 10 \end{pmatrix} & C_{11} = 2 & M_{12} = \begin{pmatrix} 4 & 6 \\ 7 & 10 \end{pmatrix} & C_{12} = 2 \\
M_{13} = \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} & C_{13} = -3 & M_{21} = \begin{pmatrix} 2 & 3 \\ 8 & 10 \end{pmatrix} & C_{21} = 4 \\
M_{22} = \begin{pmatrix} 1 & 3 \\ 7 & 10 \end{pmatrix} & C_{22} = -11 & M_{23} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix} & C_{23} = 6 \\
M_{31} = \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} & C_{31} = -3 & M_{32} = \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix} & C_{32} = 6 \\
M_{33} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} & C_{33} = -3 & &
\end{array}$$

so that:

$$\mathrm{adj}(A) = \begin{pmatrix} 2 & 2 & -3 \\ 4 & -11 & 6 \\ -3 & 6 & -3 \end{pmatrix}^T = \begin{pmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{pmatrix}.$$

The adjoint (adjunct) of a matrix is useful to us because it is related to the inverse of the matrix when this exists.

Note that if I_n is the identity matrix, then for all $i,\ j = 1, \ldots, n$, we have that:

$$(I_n)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Lemma 3

Let $A \in M_{n\times n}(\mathbb{F})$. Then

$$A\,\mathrm{adj}(A) = \mathrm{adj}(A)\,A = \det(A)I_n.$$

Proof

Let us examine the entries $(A\,\mathrm{adj}(A))_{ij}$. We can write that:

$$(A\,\mathrm{adj}(A))_{ij} = \sum_{k=1}^{n} a_{ik}(\mathrm{adj}(A))_{kj} = \sum_{k=1}^{n} a_{ik}C_{jk}.$$

(i) If $i = j$, then we have $(A\,\mathrm{adj}(A))_{ii} = \sum\limits_{k=1}^{n} a_{ik}C_{ik}$, which is none other than the i^{th} row expansion of $\det(A)$. That is,

$$(A\,\mathrm{adj}(A))_{ii} = \det(A) = \det(A)(I)_{ii}.$$

(ii) If $i \neq j$, then we will show $(A\,\mathrm{adj}(A))_{ij} = 0 = \det(A)(I)_{ij}$.

Let $B \in M_{n\times n}(\mathbb{F})$ be the matrix which you construct from A by replacing the j^{th} row of A by the i^{th} row of A. Thus the matrices A and B differ only (perhaps) in their j^{th} row, and so $C_{jk}(B) = C_{jk}(A),\ k = 1, \ldots, n$.
Since B has two identical rows, its determinant is zero.
We then evaluate $\det(B)$ by performing a jth row expansion:

$$0 = \det(B) = \sum_{k=1}^{n} b_{jk}C_{jk}(B),$$

but $b_{jk} = a_{ik}$, for $k = 1, \ldots, n$, and $C_{jk}(B) = C_{jk}(A)$. Thus

$$0 = \sum_{k=1}^{n} a_{ik}C_{jk}(A) = (A\,\mathrm{adj}(A))_{ij}.$$

We conclude that $(A\,\mathrm{adj}(A))_{ij} = \det(A)(I)_{ij}$. ■

Corollary 10

Let $A \in M_{n\times n}(\mathbb{F})$. If $\det(A) \neq 0$, then

$$A^{-1} = \left(\frac{1}{\det(A)}\right)\mathrm{adj}(A).$$

Proof

Since $A\,\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)\,I_n$, if $\det(A) \neq 0$, then

$$A\left(\frac{\mathrm{adj}(A)}{\det(A)}\right) = \left(\frac{\mathrm{adj}(A)}{\det(A)}\right)A = I_n,$$

and we conclude that

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A). \qquad \blacksquare$$

Example 15

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\det(A) \neq 0$, what is A^{-1}?

Solution

$\det(A) = ad - bc \neq 0$, thus A is invertible and

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A) = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Example 16

Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$. Find A^{-1} if it exists.

Solution

$\det(A) = \det\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{pmatrix} = 33 - 36 = -3.$

Since $\det(A) \neq 0$, then A is invertible and A^{-1} exists.

We already know from Example 14 that

$$\mathrm{adj}(A) = \begin{pmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{pmatrix}.$$

We conclude that

$$A^{-1} = \frac{1}{-3}\begin{pmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{pmatrix}.$$

Remark 3

We do not recommend this method as a way of evaluating the inverse of a matrix. However, it is very useful for checking that one or two components are correct.

Cramer's Rule

Cramer's Rule provides a method for solving systems of n linear equations in n unknowns by making use of determinants. It is not usually efficient from a computational point of view since calculating determinants can involve many calculations.

Cramer's Rule is particularly useful if you only require a single component of the solution.

Lemma 4: Cramer's Rule

Let $A \in M_{n\times n}(\mathbb{F})$. Let us consider the equation $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} \in \mathbb{F}^n$ and $\det(A) \neq 0$.

If we construct B_j from A, by replacing the j^{th} column of A by the column vector $\mathbf{b}$, then the solution $\mathbf{x}$ to the equation

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

is given by

$$x_j = \frac{\det(B_j)}{\det(A)}, \quad \text{for all } j = 1, \dots, n.$$

Proof

Since $\det(A) \neq 0$, we know that A is invertible and

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A).$$

Thus

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det A}\,\mathrm{adj}(A)\,\mathbf{b}$$

and so

$$x_j = \frac{1}{\det(A)}\sum_{k=1}^{n}(\mathrm{adj}(A))_{jk}\,b_k,$$

that is,

$$x_j = \frac{1}{\det(A)} \sum_{k=1}^{n} C_{kj}(A)\, b_k.$$

We will show that

$$\sum_{k=1}^{n} C_{kj}(A)\, b_k = \det(B_j),$$

and then Cramer's Rule will be proven.

We evaluate $\det(B_j)$ by expanding the determinant along the j^{th} column of B_j:

$$\det(B_j) = \sum_{k=1}^{n} (B_j)_{kj}\, (C(B_j))_{kj}.$$

We recall that $(B_j)_{kj} = b_k$, since the j^{th} column of B_j is just $\mathbf{b}$.

Since the matrices A and B only differ (perhaps) in the j^{th} column, we conclude that:

$$(C(B_j))_{kj} = (C(A))_{kj}.$$

And thus

$$\det(B_j) = \sum_{k=1}^{n} C_{kj}(A)\, b_k. \qquad \blacksquare$$

Example 17

Consider the matrix A from Example 16, that is :

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}.$$

Use Cramer's Rule to solve

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} \mathbf{x} = \begin{pmatrix} -2 \\ 3 \\ -4 \end{pmatrix}.$$

We already know from Example 16 that

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix} = -3.$$

We evaluate

$$\det(B_1) = \det\begin{pmatrix} -2 & 2 & 3 \\ 3 & 5 & 6 \\ -4 & 8 & 10 \end{pmatrix} = \det\begin{pmatrix} -2 & 2 & 3 \\ 0 & 8 & \frac{21}{2} \\ 0 & 4 & 4 \end{pmatrix} = 20, \qquad \begin{cases} R_2 \to R_2 + \frac{3}{2}R_1 \\ \\ R_3 \to R_3 - 2R_1 \end{cases}$$

$$\det(B_2) = \det\begin{pmatrix} 1 & -2 & 3 \\ 4 & 3 & 6 \\ 7 & -4 & 10 \end{pmatrix} = \det\begin{pmatrix} 1 & -2 & 3 \\ 0 & 11 & -6 \\ 0 & 10 & -11 \end{pmatrix} = -61, \qquad \begin{cases} R_2 \to R_2 - 4R_1 \\ \\ R_3 \to R_3 - 7R_1 \end{cases}$$

$$\det(B_3) = \det\begin{pmatrix} 1 & 2 & -2 \\ 4 & 5 & 3 \\ 7 & 8 & -4 \end{pmatrix} = \det\begin{pmatrix} 1 & 2 & -2 \\ 0 & -3 & 11 \\ 0 & -6 & 10 \end{pmatrix} = 36, \qquad \begin{cases} R_2 \to R_2 - 4R_1 \\ \\ R_3 \to R_3 - 7R_1 \end{cases}$$

Thus

$$\mathbf{x} = \frac{1}{-3}\begin{pmatrix} 20 \\ -61 \\ 36 \end{pmatrix}.$$

The Determinant and Geometry.

We begin by relating the determinant to the cross product.
Let us evaluate the cross product of the vectors $\mathbf{x} = (x_1, x_2, x_3)^T$ and $\mathbf{y} = (y_1, y_2, y_3)^T$.

$$\mathbf{x} \times \mathbf{y} = \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}.$$

Suppose we construct a "symbolic" (3×3) determinant of a matrix whose entries of its first row, are the standard basis vectors in $\mathbb{R}^3$, and whose entries in the other two rows, are the components of the two vectors $\mathbf{x}$ and $\mathbf{y}$.

Let us then evaluate this "determinant" by performing a first row expansion, obtaining thus:

$$\det\left(\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix}\right) = \mathbf{i}(1)(x_2y_3 - x_3y_2) + \mathbf{j}(-1)(x_1y_3 - x_3y_1) + \mathbf{k}(1)(x_1y_2 - x_2y_1).$$

Notice that this "symbolic" determinant is equal to the cross product of the vectors $\mathbf{x}$ and $\mathbf{y}$.

We can thus write :

$$\mathbf{x} \times \mathbf{y} = \det\left(\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix}\right).$$

Example 18

Evaluate $\mathbf{x} \times \mathbf{y}$ when $\mathbf{x} = (1, 2, 3)^T$ and $\mathbf{y} = (\,-2, 3, -4)^T$.

Solution

$$\mathbf{x} \times \mathbf{y} = \det\left(\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & 3 \\ -2 & 3 & -4 \end{pmatrix}\right).$$

Expanding along the first column, we get :

$\mathbf{x} \times \mathbf{y} = \mathbf{i}(-8-9) - \mathbf{j}(-4+6) + \mathbf{k}(3+4) = (-17, -2, 7)^T$.

We know from Topic 4 that the cross product of two vectors is related to the area of the parallelogram formed by the two vectors.

Example 19

What is the area of the parallelogram with sides $\mathbf{x} = (1, 2, 3)^T$ and $\mathbf{y} = (-2, 3, -4)^T$?

Solution

We need the length of their cross product, and so, using the result in Example 18, we find that the area of this parallelogram is:

$$\|(-17, -2, 7)^T\| = (17^2 + 2^2 + 7^2)^{1/2} = (342)^{1/2}.$$

A special case arises when you have two vectors in a plane in $\mathbb{R}^2$, and wish to know the area of the parallelogram for which the two vectors form the sides.

Lemma 5

Let $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ be vectors in $\mathbb{R}^2$. Then the **area** of the parallelogram with sides $\mathbf{v}$ and $\mathbf{w}$ is $\left|\det\left(\begin{pmatrix} v_1 & v_2 \\ w_1 & w_2 \end{pmatrix}\right)\right|$.

Proof

Let us consider the vectors as if they were in $\mathbb{R}^3$, with zero third component. That is, we let

$$\mathbf{v} = (v_1, v_2, 0)^T \quad \text{and} \quad \mathbf{w} = (w_1, w_2, 0)^T,$$

which are now vectors in $\mathbb{R}^3$, so that we can now perform their cross product.

We then obtain:

$$\mathbf{v} \times \mathbf{w} = \det\left(\begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & 0 \\ w_1 & w_2 & 0 \end{pmatrix}\right) = \mathbf{k}\det\left(\begin{pmatrix} v_1 & v_2 \\ w_1 & w_2 \end{pmatrix}\right).$$

Since the length of this vector $\mathbf{v} \times \mathbf{w}$, which has only one non-zero component, gives the area of the parallelogram, the area of the parallelogram is then the absolute value of the determinant, i.e. $\left|\det\left(\begin{pmatrix} v_1 & v_2 \\ w_1 & w_2 \end{pmatrix}\right)\right|$. $\blacksquare$

Example 20

What is the area of the parallelogram in the plane in $\mathbb{R}^2$ with the two vectors, $\mathbf{v} = (2, 5)^T$ and $\mathbf{w} = (-4, 9)^T$, as sides?

Solution

The area is $\left|\det\left(\begin{pmatrix} 2 & 5 \\ -4 & 9 \end{pmatrix}\right)\right| = 38$ units squared.

The scalar triple product of three vectors in $\mathbb{R}^3$, is defined by $STP(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \mathbf{x} \bullet (\mathbf{y} \times \mathbf{z})$. The **volume** of parallelepiped (3-d generalization of a parallelogram), which has the three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ for sides, is given by $|STP(\mathbf{x}, \mathbf{y}, \mathbf{z})|$.

Lemma 6

$$STP(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \det((\mathbf{x}, \mathbf{y}, \mathbf{z})) = \det\left((\mathbf{x}, \mathbf{y}, \mathbf{z})^T\right) = \det\left(\begin{pmatrix} \mathbf{x}^T \\ \mathbf{y}^T \\ \mathbf{z}^T \end{pmatrix}\right).$$

A simple proof is obtained by evaluating the two expressions $STP(\mathbf{x}, \mathbf{y}, \mathbf{z})$ and $\det((\mathbf{x}, \mathbf{y}, \mathbf{z}))$, and showing that they are identical.

Example 21

What is the volume of the parallelepiped with sides:

$$\mathbf{x} = (3, -5, 6)^T \ , \quad \mathbf{y} = (3, 4, 6)^T \ , \quad \mathbf{z} = (1, 5, -7)^T ?$$

Solution

We begin by evaluating $STP(\mathbf{x}, \mathbf{y}, \mathbf{z})$:

$$STP(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \det\left(\begin{pmatrix} 3 & -5 & 6 \\ 3 & 4 & 6 \\ 1 & 5 & -7 \end{pmatrix}\right) \ , \qquad R_2 \to R_2 - R_1$$

$$= \det\left(\begin{pmatrix} 3 & -5 & 6 \\ 0 & 9 & 0 \\ 1 & 5 & -7 \end{pmatrix}\right) = (9)(-21 - 6) = -243,$$

where the last determinant was evaluated with the second row expansion.
Thus, taking the absolute value, the volume of the parallelepiped is 243 (units cubed).

Example 22

For what value of the constant k are the three vectors $\mathbf{x} = (3, -5, 6)^T$,$\mathbf{y} = (3, k, 9)^T$ and $\mathbf{z} = (1, 5, -7)^T$, tangential to the same plane?

Solution

This will happen iff the volume of the parallelepiped formed by the vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ is zero.

We thus evaluate $STP(\mathbf{x}, \mathbf{y}, \mathbf{z})$:

$$STP(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \det\left(\begin{pmatrix} 3 & -5 & 6 \\ 3 & k & 9 \\ 1 & 5 & -7 \end{pmatrix}\right) \ , \qquad R_1 \leftrightarrow R_3$$

$$= -\det\left(\begin{pmatrix} 1 & 5 & -7 \\ 3 & k & 9 \\ 3 & -5 & 6 \end{pmatrix}\right) \ , \qquad \begin{cases} R_2 \to R_2 - 3R_1 \\ R_3 \to R_3 - 3R_1 \end{cases}$$

$$= -\det\left(\begin{pmatrix} 1 & 5 & -7 \\ 0 & k-15 & 30 \\ 0 & -20 & 27 \end{pmatrix}\right) \ , \qquad \text{first column expansion}$$

$$= -[(k-15)(27) - (30)(-20)] = -3\,(9k + 65).$$

We conclude that the three points are coplanar iff $k = \dfrac{-65}{9}$.

Topic 16A
Diagonalization and the Eigenvalue Problem I
Matrices

Let $A \in M_{n\times n}(\mathbb{R})$ and suppose that we consider the effect that A has on the vector $\mathbf{x} \in \mathbb{R}^n$, that is, let us compare $\mathbf{x} \in \mathbb{R}^n$ to its image, $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^n$.

There are two geometrical features which we can compare: these are the relative lengths of $\mathbf{x}$ and $\mathbf{y}$, and also the relative orientations of $\mathbf{x}$ and $\mathbf{y}$. Usually, when $\mathbf{x}$ is multiplied by matrix A to produce $\mathbf{y} = A\mathbf{x}$, there will be changes to both the length and the orientation.

Example 1

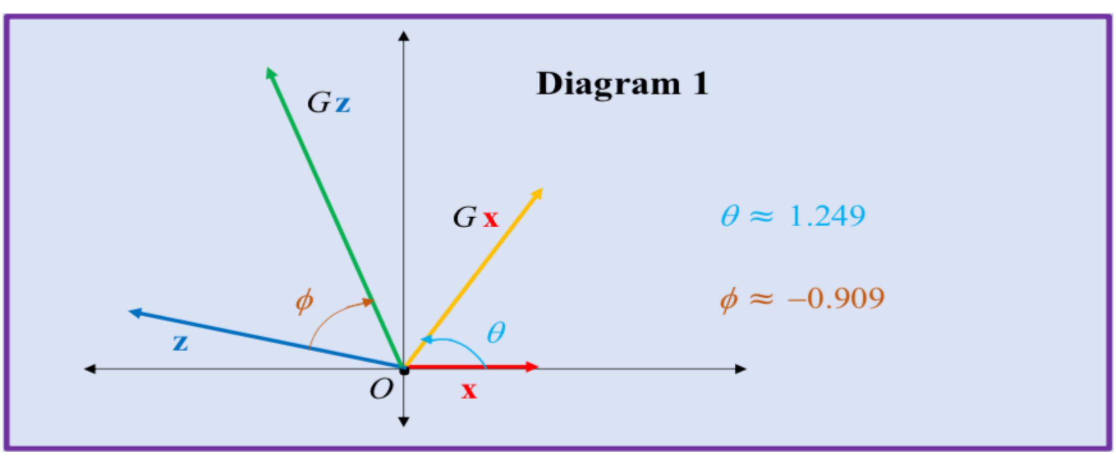
$$\text{Let } G = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \text{ and } \mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \text{ Then } \mathbf{y} = T_G(\mathbf{x}) = G\,\mathbf{x} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

The length of the image, $\mathbf{y}$, is $\|\mathbf{y}\| = \sqrt{10}$, that is, the image had been scaled by a factor of $\sqrt{10}$ relative to the length of the original vector $\mathbf{x}$, $\|\mathbf{x}\| = 1$.

In addition, the image makes an angle $\theta = \arctan(3) = 1.249$ radians with the positive x-axis, while the original vector, $\mathbf{x}$, makes an angle of 0 radians with the positive x-axis: this corresponds to an **anti-clockwise** rotation of 1.249 radians.

The image of the vector $\mathbf{z} = (2, -1)^T$ is the vector $\mathbf{w} = G\,\mathbf{z} = (-1, 5)^T$. Since $\|\mathbf{z}\| = \sqrt{5}$ and $\|\mathbf{w}\| = \sqrt{26}$, the vector $\mathbf{z}$ has been scaled by a factor of $\sqrt{\frac{26}{5}}$ by the function T_G.

In addition, the vector $\mathbf{z}$ makes an angle of $\arctan(-1/2)$ with the positive x-axis, and the vector $\mathbf{w}$ makes an angle of $\arctan(-5)$ with the positive x-axis, so that $\mathbf{z}$ has been rotated by the function T_G, by an angle $\phi = [\arctan(-5) - \arctan(-1/2)] = 0.909$ radians **clockwise**.



In this section of the course, we ask the question, for a given matrix $A \in M_{n\times n}(\mathbb{F})$, are there any vectors that are just scaled when they are multiplied by A.

Example 1 - continued

Consider the matrix G in Example 1 and the vector $(1,1)^T$. We then have:

$$T_G\left(\begin{pmatrix}1\\1\end{pmatrix}\right) = G\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1 & 3\\3 & 1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}4\\4\end{pmatrix} = 4\begin{pmatrix}1\\1\end{pmatrix}.$$

Thus the vector $(1,1)^T$ is scaled by 4 by the matrix G, and it is not rotated by multiplying it by this matrix.

Note that we include, within our scope of interest, vectors which have a negative scaling factor. Although you could argue that they are rotated by an angle of π.

Example 1 - continued

Consider the matrix G in Example 1 and the vector $(1,-1)^T$. We now have:

$$T_G\left(\begin{pmatrix}1\\-1\end{pmatrix}\right) = G\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}1 & 3\\3 & 1\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}-2\\2\end{pmatrix} = -2\begin{pmatrix}1\\-1\end{pmatrix}.$$

We consider the vector $(1,-1)^T$ to be scaled by a factor of (-2) by the matrix G, instead of saying that it is scaled by a factor of (2) and rotated anti-clockwise by an angle π.

Also note that since $G\,\mathbf{0} = \mathbf{0} = k\,\mathbf{0}$, for any constant k: we are thus not particularly interested in the zero vector.

Although we have motivated this problem geometrically, and therefore we have been considering $\mathbb{R}^n$, we will be approaching this question in both $\mathbb{R}^n$ and $\mathbb{C}^n$.

We will see that we will now need to be careful to remember which of these fields we are using.

The mathematical formulation of the problem is as follows:

Definition 1: Eigenvector, eigenvalue and eigenpair.

Let $A \in M_{n\times n}(\mathbb{F})$. We then say that the **non-zero** vector $\mathbf{x}$ is an **eigenvector** of A to mean that there exists a scalar $\lambda \in \mathbb{F}$ such that

$$A\,\mathbf{x} = \lambda\,\mathbf{x}.$$

The scalar λ is called an **eigenvalue,** and the pair $(\lambda, \mathbf{x})$ is called an **eigenpair**.

Let $A \in M_{n\times n}(\mathbb{F})$ and I_n be denoted by simply I.
Solving the matrix eigenvalue problem is solving:

$$\begin{aligned} & A\mathbf{x} = \lambda\mathbf{x} \\ & \quad \Leftrightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \\ & \quad \Leftrightarrow A\mathbf{x} - \lambda I\mathbf{x} = \mathbf{0} \\ & \quad \Leftrightarrow (A - \lambda I)\mathbf{x} = \mathbf{0} \\ & \quad \Leftrightarrow M\mathbf{x} = \mathbf{0} \quad \text{where } M = (A - \lambda I). \end{aligned}$$

Solving the equation $A\mathbf{x} = \lambda\mathbf{x}$ is thus equivalent to solving $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

Definition 2: Eigenvalue equation or Eigenvalue problem

We refer to the equation

$$A\,\mathbf{x} = \lambda\,\mathbf{x} \quad \text{or} \quad (A - \lambda I)\,\mathbf{x} = \mathbf{0}$$

as the **eigenvalue equation**, or the **eigenvalue problem**.

This is an unusual problem since we do not know the vector $\mathbf{x} \in \mathbb{F}^n$, nor the scalar $\lambda \in \mathbb{F}$.
We want to obtain a non-trivial ($\mathbf{x} \neq \mathbf{0}$) solution to the eigenvalue equation.
This is possible **iff** the matrix M is not invertible.
This is true **iff** $\det(M) = 0$.
Thus, if we can find a special number λ such that $\det(A - \lambda I) = 0$, then we have found an eigenvalue.
If we then examine the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$, then we will be guaranteed to be able to find an eigenvector $\mathbf{x}$: i.e. to find a non-trivial solution to the eigenvalue equation.

So how do we find an eigenvalue?

Definition 3: Characteristic polynomial and characteristic equation.

Let $A \in M_{n\times n}(\mathbb{F})$ and $t \in \mathbb{F}$.
We consider the expression :

$$\det(A - tI).$$

The **characteristic polynomial** of A means the polynomial, $\Delta_A(t)$, defined by,

$$\Delta_A(t) = \det(A - tI).$$

The **characteristic equation** of A is the equation

$$\Delta_A(t) = \det(A - tI) = 0.$$

The eigenvalues of A are thus the roots of the characteristic polynomial of A.
We will usually label the eigenvalues λ_i of A in **decreasing order** of their numerical value.

Once you have found an eigenvalue, then you can examine the eigenvalue equation in the form $(A - \lambda I)\mathbf{x} = \mathbf{0}$, in order to obtain the corresponding eigenvector.

Example 2

Solve the eigenvalue problem for

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix},$$

where we consider $A \in M_{2\times 2}(\mathbb{R})$.

Solution

(i) $A - tI = \begin{pmatrix} 1-t & 2 \\ 2 & 1-t \end{pmatrix}$, so that the characteristic polynomial of A is:

$$\Delta_A(t) = (1-t)^2 - 2^2 = (3-t)(-1-t).$$

and the characteristic equation of A is

$$\Delta_A(t) = (1-t)^2 - 2^2 = (3-t)(-1-t) = 0.$$

(ii) The zeros of $\Delta_A(t)$ are the eigenvalues of A, thus $\lambda_1 = 3$ and $\lambda_2 = -1$.

(iii) For each eigenvalue, λ_i, we examine the eigenvalue equation to determine the eigenvectors.

(I) $\lambda_1 = 3$. We then examine

$$(A - 3I)\,\mathbf{x} = \mathbf{0}.$$

That is,

$$\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{0}\ , \quad \text{where } \mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2.$$

Simplification yields

$$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{0}.$$

The general solution is

$$\left\{ s\begin{pmatrix} 1 \\ 1 \end{pmatrix} : s \in \mathbb{R} \right\}.$$

An eigenpair is

$$\left(3, \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right).$$

(II) $\lambda_2 = -1$. We then examine

$$[A - (-1)I]\,\mathbf{x} = \mathbf{0}.$$

That is

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} c \\ d \end{pmatrix} = \mathbf{0}\ , \quad \text{where } \mathbf{x} = \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{R}^2.$$

Simplification yields

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \mathbf{0}.$$

The general solution is

$$\left\{ t \begin{pmatrix} 1 \\ -1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

An eigenpair is

$$\left(-1, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right).$$

Example 3

Solve the eigenvalue problem for

$$B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

where we consider $B \in M_{2\times 2}(\mathbb{R})$.

Solution

(i) $B - tI = \begin{pmatrix} -t & 1 \\ -1 & -t \end{pmatrix}$, so that the characteristic polynomial of B is:

$$\Delta_B(t) = t^2 + 1$$

and the characteristic equation of B is

$$\Delta_B(t) = t^2 + 1 = 0.$$

Since $\Delta_B(t)$ has no real roots, B has no eigenvalues and no eigenpairs.

Example 4

Solve the eigenvalue problem for

$$C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

where we consider $C \in M_{2\times 2}(\mathbb{C})$.

Solution

(i) $C - tI = \begin{pmatrix} -t & 1 \\ -1 & -t \end{pmatrix}$, so that the characteristic polynomial of C is:

$$\Delta_C(t) = t^2 + 1 = (t+i)(t-i)$$

and the characteristic equation of C is:

$$\Delta_C(t) = t^2 + 1 = (t+i)(t-i) = 0.$$

(ii) The zeros of $\Delta_C(t)$ are the eigenvalues of C, thus $\lambda_1 = i,\ \lambda_2 = -i$.

(iii) For each eigenvalue, λ_i, we examine the eigenvalue equation to determine the eigenvectors.

(I) $\lambda_1 = i$. We examine

$$(C - iI)\mathbf{x} = \mathbf{0}.$$

That is,

$$\begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \mathbf{0}\ , \quad \text{where } \mathbf{x} = \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{C}^2.$$

Simplification yields

$$\begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \mathbf{0}.$$

The general solution is

$$\left\{ s \begin{pmatrix} i \\ -1 \end{pmatrix} : s \in \mathbb{C} \right\}.$$

An eigenpair is

$$\left(i, \begin{pmatrix} i \\ -1 \end{pmatrix} \right).$$

(II) $\lambda_2 = -i$. We examine

$$(C - (-i)I)\mathbf{x} = \mathbf{0},$$

that is,

$$\begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{0}\ , \quad \text{where } \mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{C}^2.$$

Simplification yields

$$\begin{pmatrix} i & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{0}.$$

The general solution is

$$\left\{ t \begin{pmatrix} 1 \\ -i \end{pmatrix} : t \in \mathbb{C} \right\}.$$

An eigenpair is

$$\left(-i, \begin{pmatrix} 1 \\ -i \end{pmatrix} \right).$$

Example 5

Solve the eigenvalue problem for

$$D = \begin{pmatrix} 3i & -4 \\ 2 & i \end{pmatrix}.$$

where we consider $D \in M_{2\times 2}(\mathbb{C})$.

Solution

(i) $D - tI = \begin{pmatrix} 3i - t & -4 \\ 2 & i - t \end{pmatrix}$ so that the characteristic polynomial is:

$$\Delta_D(t) = (3i - t)(i - t) + 8 = 3i^2 + t^2 - 4it + 8 = t^2 - 4it + 5,$$

and the characteristic equation is:

$$\Delta_D(t) = t^2 - 4it + 5 = 0.$$

(ii) The zeros of $\Delta_D(t)$ are:

$$\frac{4i \pm (-16 - 20)^{1/2}}{2} = 2i \pm 3i$$

so that the eigenvalues are $\lambda_1 = 5i$ and $\lambda_2 = -i$.

(iii) We now examine the eigenvalue equation for each of these two eigenvalues.

(I) $\lambda_1 = 5i$. We examine

$$(D - 5iI)\mathbf{x} = \mathbf{0}$$

becomes

$$\begin{pmatrix} -2i & -4 \\ 2 & -4i \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \mathbf{0}\ , \quad \text{where } \mathbf{x} = \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{C}^2.$$

The general solution is

$$\left\{ s \begin{pmatrix} 2i \\ 1 \end{pmatrix} : s \in \mathbb{C} \right\}.$$

An eigenpair is

$$\left(5i, \begin{pmatrix} 2i \\ 1 \end{pmatrix} \right).$$

(II) $\lambda_2 = -i$.

$$(D - (-i)I)\mathbf{x} = \mathbf{0}$$

That is,

$$\begin{pmatrix} 4i & -4 \\ 2 & 2i \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{0}\ , \quad \text{where } \mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{C}^2.$$

Simplification yields

$$\begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{0}.$$

The general solution is

$$\left\{ t \begin{pmatrix} -i \\ 1 \end{pmatrix} : t \in \mathbb{C} \right\}.$$

An eigenpair is

$$\left(-i, \begin{pmatrix} -i \\ 1 \end{pmatrix} \right).$$

Example 6

Solve the eigenvalue problem for the matrix

$$F = \begin{pmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{pmatrix}.$$

Solution

We first evaluate the characteristic polynomial of F, namely $\Delta_F(t)$:

$$\begin{aligned}
\Delta_F(t) = \det &\left(\begin{pmatrix} 4-t & 2 & -6 \\ 1 & -2-t & 1 \\ -6 & 2 & 4-t \end{pmatrix} \right) \\
&= (4-t)((-2-t)(4-t)-2) - 2(1(4-t)-1(-6)) - 6(1(2)-(-2-t)(-6)) \\
&= -(4-t)^2(2+t) - 8 + 2t - 20 + 2t + 60 + 36t = -(4-t)^2(2+t) + 32 + 40t \\
&= -t^3 + 6t^2 + 40t = -t(t^2 - 6t - 40) = t(t-10)(t+4).
\end{aligned}$$

We then solve the characteristic equation $\Delta_F(t) = t(t-10)(t+4) = 0$.

Thus the eigenvalues of F are $\lambda_1 = 10, \lambda_2 = 0, \lambda_3 = -4$.

We now find their corresponding eigenvectors.

For $\lambda_1 = 10$, we consider the eigenvalue equation:

$$(F - \lambda_1 I)\mathbf{x}_1 = \mathbf{0}\ , \quad \text{which is } \begin{pmatrix} -6 & 2 & -6 \\ 1 & -12 & 1 \\ -6 & 2 & -6 \end{pmatrix} \mathbf{x}_1 = \mathbf{0},$$

and we consider the augmented matrix:

$$\left(\begin{array}{rrr|r} -6 & 2 & -6 & 0 \\ 1 & -12 & 1 & 0 \\ -6 & 2 & -6 & 0 \end{array} \right), \text{ which we row reduce.}$$

$$R_1 \to \frac{-1}{6}R_1 \qquad \left(\begin{array}{rrr|r} 1 & \frac{-1}{3} & 1 & 0 \\ 1 & -12 & 1 & 0 \\ -6 & 2 & -6 & 0 \end{array}\right) \qquad \xrightarrow[R_3 \to R_3+6R_1]{R_2 \to R_2 - R_1} \qquad \left(\begin{array}{rrr|r} 1 & \frac{-1}{3} & 1 & 0 \\ 0 & \frac{-35}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$$R_2 \to \frac{-3}{35}R_2 \qquad \left(\begin{array}{rrr|r} 1 & \frac{-1}{3} & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \qquad R_1 \to R_1 + \tfrac{1}{3}R_2 \qquad \left(\begin{array}{rrr|r} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The solution set is $\left\{s\,(1,0,-1)^T : s \in \mathbb{R}\right\}$, and an eigenpair is $\left(10,(1,0,-1)^T\right)$.

For $\lambda_2 = 0$, we consider the eigenvalue equation:

$$(F - \lambda_2 I)\mathbf{x}_2 = \mathbf{0}\ , \quad \text{which is} \quad \begin{pmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{pmatrix}\mathbf{x}_2 = \mathbf{0},$$

and we consider the augmented matrix:

$$\left(\begin{array}{rrr|r} 4 & 2 & -6 & 0 \\ 1 & -2 & 1 & 0 \\ -6 & 2 & 4 & 0 \end{array}\right), \text{ which we row reduce.}$$

$$R_1 \to \frac{1}{4}R_1 \qquad \left(\begin{array}{rrr|r} 1 & \frac{1}{2} & \frac{-3}{2} & 0 \\ 1 & -2 & 1 & 0 \\ -6 & 2 & 4 & 0 \end{array}\right) \qquad \xrightarrow[R_3 \to R_3+6R_1]{R_2 \to R_2 - R_1} \qquad \left(\begin{array}{rrr|r} 1 & \frac{1}{2} & \frac{-3}{2} & 0 \\ 0 & \frac{-5}{2} & \frac{5}{2} & 0 \\ 0 & 5 & -5 & 0 \end{array}\right)$$

$$R_2 \to \frac{-2}{5}R_2 \qquad \left(\begin{array}{rrr|r} 1 & \frac{1}{2} & \frac{-3}{2} & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 5 & -5 & 0 \end{array}\right) \qquad \xrightarrow[R_3 \to R_3-5R_2]{R_1 \to R_1 - \frac{1}{2}R_2} \qquad \left(\begin{array}{rrr|r} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

The solution set is $\left\{s\,(1,1,1)^T : s \in \mathbb{R}\right\}$, and an eigenpair is $\left(0,(1,1,1)^T\right)$.

For $\lambda_3 = -4$, we consider the eigenvalue equation:

$$(F - \lambda_3 I)\mathbf{x}_3 = \mathbf{0}\ , \quad \text{which is} \quad \begin{pmatrix} 8 & 2 & -6 \\ 1 & 2 & 1 \\ -6 & 2 & 8 \end{pmatrix}\mathbf{x}_3 = \mathbf{0},$$

and we consider the augmented matrix:

$$\left(\begin{array}{rrr|r} 8 & 2 & -6 & 0 \\ 1 & 2 & 1 & 0 \\ -6 & 2 & 8 & 0 \end{array}\right), \text{ which we row reduce.}$$

$$R_1 \to \frac{1}{8}R_1 \qquad \left(\begin{array}{rrrl} 1 & \frac{1}{4} & \frac{-3}{4} & \mid 0 \\ 1 & 2 & 1 & \mid 0 \\ -6 & 2 & 8 & \mid 0 \end{array}\right) \xrightarrow[R_3 \to R_3 + 6R_1]{R_2 \to R_2 - R_1} \left(\begin{array}{rrrl} 1 & \frac{1}{4} & \frac{-3}{4} & \mid 0 \\ 0 & \frac{7}{4} & \frac{7}{4} & \mid 0 \\ 0 & \frac{7}{2} & \frac{7}{2} & \mid 0 \end{array}\right)$$

$$R_2 \to \frac{4}{7}R_2 \qquad \left(\begin{array}{rrrl} 1 & \frac{1}{4} & \frac{-3}{4} & \mid 0 \\ 0 & 1 & 1 & \mid 0 \\ 0 & \frac{7}{2} & \frac{7}{2} & \mid 0 \end{array}\right) \xrightarrow[R_3 \to R_3 - \frac{7}{2}R_2]{R_1 \to R_1 - \frac{1}{2}R_2} \left(\begin{array}{rrrl} 1 & 0 & -1 & \mid 0 \\ 0 & 1 & 1 & \mid 0 \\ 0 & 0 & 0 & \mid 0 \end{array}\right)$$

The solution set is $\left\{ s\,(1,-1,1)^T : s \in \mathbb{R} \right\}$, and an eigenpair is $\left(-4, (1,-1,1)^T\right)$.

Remark 1

If $(\lambda_1, \mathbf{x})$ is an eigenpair of an $n \times n$ matrix A, and if $k \in \mathbb{F} \setminus \{\mathbf{0}\}$, then $(\lambda_1, k\mathbf{x})$ is also an eigenpair of the matrix A. It seems that it would be a good idea to collect all these eigenvectors together, and that is what we do with the next definition.

Definition 4: Eigenspace

Let $A \in M_{n\times n}(\mathbb{F})$, and suppose that λ_1 is an eigenvalue of A. We then define the **eigenspace** of A (associated with the eigenvalue λ_1), E_{λ_1}, to be the solution set to the system of equations, $(A - \lambda_1 I)\mathbf{x} = \mathbf{0}$. That is,

$$E_{\lambda_1} = N(A - \lambda_1\, I).$$

Notice that if E_{λ_1} is an eigenspace of the matrix A, then E_{λ_1} contains all of the eigenvectors of A which have eigenvalue λ_1, **together with the zero vector**, since the eigenvalue equation has always the trivial solution.

Example 7

Let us refer back to Examples 2, 5, 6, already seen in this topic.

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix},\ E_{-1} = Span\left(\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}\right),\ E_3 = Span\left(\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}\right).$$

$$D = \begin{pmatrix} 3i & -4 \\ 2 & i \end{pmatrix},\ E_{5i} = Span\left(\left\{\begin{pmatrix} 2i \\ 1 \end{pmatrix}\right\}\right),\ E_{-i} = Span\left(\left\{\begin{pmatrix} -i \\ 1 \end{pmatrix}\right\}\right).$$

$$F = \begin{pmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{pmatrix},\ E_{10} = Span\left(\left\{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\right\}\right),$$

$$E_0 = Span\left(\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}\right),\ E_{-4} = Span\left(\left\{\begin{pmatrix}1\\-1\\1\end{pmatrix}\right\}\right).$$

Example 8

Find the eigenspaces of the matrix $Z = \begin{pmatrix}3 & 1 & 1\\1 & 3 & 1\\1 & 1 & 3\end{pmatrix}$.

Solution

We begin by examining the characteristic polynomial, $\Delta_Z(t)$.

$$\Delta_Z(t) = \det\left(\begin{pmatrix}3-t & 1 & 1\\1 & 3-t & 1\\1 & 1 & 3-t\end{pmatrix}\right)$$

$$= (3-t)((3-t)^2-1) - 1((3-t)-1) + 1(1-(3-t))$$

$$= (3-t)^3 + 3t - 7 = -t^3 + 9t^2 - 27t + 27 + 3t - 7$$

$$= -t^3 + 9t^2 - 24t + 20 = -(t-5)(t-2)^2.$$

Thus the eigenvalues of Z are $\lambda_1 = 5$, and $\lambda_2 = 2$.

We now evaluate their eigenspaces.

For $\lambda_1 = 5$, we obtain E_5 by solving:

$$(Z-5I)\mathbf{x} = \mathbf{0},\quad \text{which is} \quad \begin{pmatrix}-2 & 1 & 1\\1 & -2 & 1\\1 & 1 & -2\end{pmatrix}\mathbf{x} = \mathbf{0}.$$

After row reducing, an equivalent system is found to be:

$$\begin{pmatrix}1 & 0 & -1\\0 & 1 & -1\\0 & 0 & 0\end{pmatrix}\mathbf{x} = \mathbf{0}.$$

We can let $x_3 = u \in \mathbb{R}$, and then $x_2 = u$ and $x_1 = u$. So that

$$E_5 = \left\{u\begin{pmatrix}1\\1\\1\end{pmatrix} : u \in \mathbb{R}\right\} = Span\left(\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}\right).$$

For $\lambda_2 = 2$, we obtain E_2 by solving:

$$(Z - 2I)\mathbf{x} = \mathbf{0}, \quad \text{which is} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \mathbf{x} = \mathbf{0}.$$

After row reducing, an equivalent system is found to be:

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0}.$$

We can let $x_3 = t \in \mathbb{R}$, and $x_2 = s \in \mathbb{R}$, and then $x_1 = -s - t$. So that

$$E_2 = \left\{ s \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} : s, t \in \mathbb{R} \right\} = Span \left(\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} \right).$$

Topic 16B

Diagonalization and the Eigenvalue Problem II

Definition 5: Similar

Let $A,\ B \in M_{n\times n}(\mathbb{F})$. We say that A is **similar** to B to mean that there exists an invertible matrix $Q \in M_{n\times n}(\mathbb{F})$ such that $Q^{-1}AQ = B$.

Definition 6

The transformation, $T : M_{n\times n}(\mathbb{F}) \to M_{n\times n}(\mathbb{F})$, defined by $T(A) = Q^{-1}AQ$, or simply, $A \to Q^{-1}AQ$, is called a **similarity transformation.**

Example 9

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \text{if } Q = \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix}, \quad \text{then}$$

$$\begin{aligned} Q^{-1}AQ &= \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix} \\ &= \tfrac{-1}{2} \begin{pmatrix} 6 & -2 \\ -4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix}, \text{ is a similarity transformation,} \\ Q^{-1}AQ &= \tfrac{1}{2} \begin{pmatrix} -16 & -24 \\ 17 & 26 \end{pmatrix} = B, \text{ say.} \end{aligned}$$

The matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ is similar to the matrix $B = \frac{1}{2} \begin{pmatrix} -16 & -24 \\ 17 & 26 \end{pmatrix}$.

Notice that if $Q^{-1}AQ = B$, then $(Q^{-1})^{-1}BQ^{-1} = A$, so that if A is similar to B, B is similar to A: we just say that A and B are similar (to each other).

Definition 7: Trace

Let $A \in M_{n\times n}(\mathbb{F})$. We define the **trace** of $A, tr(A)$, to mean

$$tr(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} (A)_{ii}.$$

That is, the trace of a square matrix is the sum of its diagonal entries.

Lemma 1

If A and B are similar, then

$$(i)\ \det(A) = \det(B)$$

and

$$(ii)\ tr(A) = tr(B).$$

Proof

Recall that we may use both notation $(C)_{st}$ or c_{st} to refer to the $(s,t)^{th}$ entry of a matrix C.

Let A and B be similar. Then there exists an invertible Q such that $Q^{-1}AQ = B$.
Using Corollaries 8 and 9 in Topic 15B for part (i), we have that:

$$\begin{aligned}(i)\ \det(B) &= \det(Q^{-1}AQ) = \det(Q^{-1})\det(A)\det(Q)\\ &= (\det(Q))^{-1}\det(Q)\det(A) = \det(A).\end{aligned}$$

$$\begin{aligned}(ii)\ tr(B) &= \sum_{i=1}^{n} b_{ii} = \sum_{i=1}^{n}(Q^{-1}AQ)_{ii} = \sum_{i=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n}(Q^{-1})_{is}(A)_{st}(Q)_{ti}\\ &= \sum_{i=1}^{n}\sum_{t=1}^{n}\frac{\sum_{s=1}^{n} C_{si}(Q)}{\det(Q)}a_{st}(Q)_{ti} = \frac{1}{\det(Q)}\sum_{s=1}^{n}\sum_{t=1}^{n}a_{st}\sum_{i=1}^{n}(Q)_{ti}C_{si}(Q)\\ &= \frac{1}{\det(Q)}\sum_{s=1}^{n}\sum_{t=1}^{n}a_{st}\det(Q)(I)_{ts} = \sum_{s=1}^{n}a_{ss} = tr(A).\end{aligned}$$

The second last step in part (ii) uses Lemma 3 of Topic 15C, i.e. $Q\,\mathrm{adj}(Q) = \det(Q)I_n$. ■

In practice, matrices with lots of zeros in them are easier to deal with than matrices which do not have lots of zeros in them. In particular, matrices which are diagonal are very simple to manipulate. When we do not have a diagonal matrix, we can ask if it is possible to "convert" our matrix into one that is diagonal.

Definition 8: Diagonalizable matrix

Let $A \in M_{n\times n}(\mathbb{F})$. We say that A is **diagonalizable** to mean that there exists an invertible matrix $P \in M_{n\times n}(\mathbb{F})$ such that $P^{-1}AP = D$, where D is a diagonal matrix.

That is, A is **diagonalizable** means that A is similar to a diagonal matrix.

Example 10

The matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ is diagonalizable since

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{-1}{2} \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 3 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}, \quad \text{which is diagonal.}$$

There is a very close connection between the eigenvalue problem for a matrix and the question of whether or not the matrix is diagonalizable.

The following result illustrates that connection in a particular situation.

Lemma 2: Diagonalization I - the simplest case.

Let $A \in M_{n\times n}(\mathbb{F})$ have eigenpairs $(\lambda_1, \mathbf{v}_1), (\lambda_2, \mathbf{v}_2), \ldots, (\lambda_n, \mathbf{v}_n)$, where the eigenvalues of A, $\lambda_1, \lambda_2, \ldots, \lambda_n$, are all different.

Let $P = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. Then,

$$P \text{ is invertible and } P^{-1}AP = D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

That is, if the matrix A has n different eigenvalues, then A is diagonalizable.
Moreover, the matrix, P, which diagonalizes A, is just the matrix of the eigenvectors of A, and when it is diagonalized, the resulting diagonal matrix, D, is the diagonal matrix whose entries are the eigenvalues of A.

We will have more general versions of this result later on, and we will prove them all at the same time.

We will now apply Lemma 2 to the examples from Topic 16A, where appropriate.

Example 11

(i) Example 2 in Topic 16A revisited.

The matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ is diagonalizable. If we construct the matrix P from the eigenvectors obtained in Example 2 in Topic 16A: we then have, $P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and a quick calculation gives that $P^{-1} = \frac{-1}{2} \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}$.

Then, according to Lemma 2, $P^{-1}AP = diag(3, -1)$. Let us check this:

$$P^{-1}AP = \tfrac{-1}{2}\begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$
$$= \tfrac{-1}{2}\begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 3 & -1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix},$$

and we have verified Lemma 2.

(ii) Example 4 in Topic 16A revisited.

The matrix $C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is diagonalizable. If we construct the matrix P from the eigenvectors obtained in Example 4 in Topic 16A: we then have, $P = \begin{pmatrix} i & 1 \\ -1 & -i \end{pmatrix}$, and a quick calculation gives that $P^{-1} = \frac{1}{2}\begin{pmatrix} -i & -1 \\ 1 & i \end{pmatrix}$.

Then, according to Lemma 2, $P^{-1}CP = diag(i, -i)$. Let us check this:

$$P^{-1}CP = \tfrac{1}{2}\begin{pmatrix} -i & -1 \\ 1 & i \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} i & 1 \\ -1 & -i \end{pmatrix}$$
$$= \tfrac{1}{2}\begin{pmatrix} -i & -1 \\ 1 & i \end{pmatrix}\begin{pmatrix} -1 & -i \\ -i & -1 \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

and we have verified Lemma 2.

(iii) Example 5 in Topic 16A revisited.

The matrix $D = \begin{pmatrix} 3i & -4 \\ 2 & i \end{pmatrix}$ is diagonalizable. If we construct the matrix P from the eigenvectors obtained in Example 5 in Topic 16A: we then have, $P = \begin{pmatrix} 2i & -i \\ 1 & 1 \end{pmatrix}$, and a quick calculation gives that $P^{-1} = \frac{1}{3i}\begin{pmatrix} 1 & i \\ -1 & 2i \end{pmatrix}$.

Then, according to Lemma 2, $P^{-1}DP = diag(5i, -i)$. Let us check this:

$$P^{-1}DP = \tfrac{-i}{3}\begin{pmatrix} 1 & i \\ -1 & 2i \end{pmatrix}\begin{pmatrix} 3i & -4 \\ 2 & i \end{pmatrix}\begin{pmatrix} 2i & -i \\ 1 & 1 \end{pmatrix}$$
$$= \tfrac{-i}{3}\begin{pmatrix} 1 & i \\ -1 & 2i \end{pmatrix}\begin{pmatrix} -10 & -1 \\ 5i & -i \end{pmatrix} = \begin{pmatrix} 5i & 0 \\ 0 & -i \end{pmatrix},$$

and we have verified Lemma 2.

(iv) Example 6 in Topic 16A revisited.

The matrix $F = \begin{pmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{pmatrix}$ is diagonalizable. If we construct the matrix P from the eigenvectors obtained in Example 6 in Topic 16A: we then have, $P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 1 & 1 \end{pmatrix}$,

and a routine calculation gives that $P^{-1} = \frac{1}{4} \begin{pmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{pmatrix}$.

Then, according to Lemma 2, $P^{-1}FP = diag(10, 0, -4)$. Let us check this:

$$\begin{aligned} P^{-1}FP &= \tfrac{1}{4} \begin{pmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 1 & 1 \end{pmatrix} \\ &= \tfrac{1}{4} \begin{pmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 10 & 0 & -4 \\ 0 & 0 & 4 \\ -10 & 0 & -4 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -4 \end{pmatrix}, \end{aligned}$$

and we have verified Lemma 2.

Lemma 3: Properties of the characteristic polynomial.

Let $A \in M_{n\times n}(\mathbb{F})$ have characteristic polynomial $\Delta_A(t) = \det(A - tI)$. We then have:

(i) $\Delta_A(t)$ is an n^{th} order polynomial in t, and so we may write it as

$$\Delta_A(t) = b_0 + b_1 t + \cdots + b_{(n-1)} t^{(n-1)} + b_n t^n.$$

(ii) $b_n = (-1)^n$.

(iii) $b_{(n-1)} = (-1)^{(n-1)} tr(A)$.

(iv) $b_0 = \det(A)$.

Outline of the proof

The best way to prove this result is to use the so-called classical definition of the determinant, which some of you will meet in other courses.
Here we just make some observations.

When you expand the determinant, using the first row expansion (Topic 15A), then you obtain a sum of $n!$ terms, each of which is the product of n terms, one from each row, and also one from each column.

One of the terms in the expansion of the determinant of $(A - tI)$ is the product of the diagonal entries,
$$(a_{11} - t)(a_{22} - t) \dots (a_{(n-1)(n-1)} - t)(a_{nn} - t).$$
This is the only term in the expansion of the determinant of $(A - tI)$ which involves n linear terms in t. Any other term in the expansion of the determinant of $(A - tI)$ will involve at most $(n - 2)$ terms that are linear in t, and the other terms will be constants.
Hence, all the other terms will be of order at most $(n - 2)$ in t.
Thus the characteristic polynomial of A will be a polynomial of order at most n.

The term $(a_{11} - t)(a_{22} - t) \dots (a_{(n-1)(n-1)} - t)(a_{nn} - t)$, provides a term in t^n with coefficient $(-1)^n$, and a term in $t^{(n-1)}$ with coefficient $(-1)^{(n-1)}(a_{11} + \dots + a_{nn})$.
So we can conclude that:

(i) $\Delta_A(t)$ is an n^{th} order polynomial in t, which we write as

$$\Delta_A(t) = b_0 + b_1 t + \dots + b_{(n-1)} t^{(n-1)} + b_n t^n,$$

(ii) $b_n = (-1)^n$,

(iii) $b_{(n-1)} = (-1)^{(n-1)}(a_{11} + \dots + a_{nn}) = (-1)^{(n-1)} tr(A)$.

Notice that if we put $t = 0$, into the characteristic polynomial we get:

$$\Delta_A(0) = \det(A - 0I) = \det(A) = b_0,$$

and thus the constant term in the characteristic polynomial of A is actually the determinant of A, and so part (iv) is true too. ■

The reason why Lemma 3 is useful, is that there is a lot of arithmetic involved in solving the eigenvalue problem, and you want to make sure that your characteristic polynomial is actually correct before you perform too many additional steps. Since it is easy to obtain the trace of the original matrix, and its determinant can usually be obtained relatively quickly, this lemma provides FOUR quick checks to perform before you start trying to factor the characteristic polynomial and find the eigenvalues.

Example 12

What is the characteristic polynomial of $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{pmatrix}$?

Note that we have already evaluated $\det(A) = -3$, for this matrix, and notice that $n = 3$, so that only one term will provide cubic and quadratic terms in t.

Solution

$$\begin{aligned}\Delta_A(t) &= \det\left(\begin{pmatrix} 1-t & 2 & 3 \\ 4 & 5-t & 6 \\ 7 & 8 & 10-t \end{pmatrix}\right)\\ &= (1-t)\left[(5-t)(10-t)-48\right] - 2\left[4(10-t)-42\right] + 3\left[32-7(5-t)\right]\\ &= (1-t)(5-t)(10-t) - 48 + 48t + 4 + 8t - 9 + 21t\\ &= -t^3 + 16t^2 - 65t + 50 - 53 + 77t = -t^3 + 16t^2 + 12t - 3.\end{aligned}$$

If at this point, we were a little worried about whether or not this expression is correct, then we observe that:

(i) and (ii) It is a cubic polynomial with coefficient of t^3 being $(-1) = (-1)^3$, and

(iii) The coefficient of t^2 is $(16) = (-1)^2\, tr(A)$, and

(iv) The constant term is $-3 = \det(A)$.

We should now be rather confident that our characteristic polynomial is correct.

Since the characteristic polynomial is an n^{th} order polynomial, then there will be at most n eigenvalues for an $n \times n$ matrix, and in fact there will be exactly n eigenvalues, if we are working over the complex field. Note the n eigenvalues need not be distinct.

The following result is useful for checking the eigenvalues, when we are working over the complex field.

Lemma 4: Properties of the characteristic polynomial.

Let $A \in M_{n\times n}(\mathbb{C})$ with characteristic polynomial

$$\Delta_A(t) = b_0 + b_1 t + \cdots + b_{(n-1)} t^{(n-1)} + b_n t^n,$$

and suppose that A has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then

(i) $\sum_{i=1}^{n} \lambda_i = tr(A) = (-1)^{(n-1)} b_{(n-1)}$

and

(ii) $\prod_{i=1}^{n} \lambda_i = \det(A) = b_0.$

Proof

The eigenvalues are the roots of the characteristic polynomial, and so we can write:

$$\Delta_A(t) = c(\lambda_1 - t)(\lambda_2 - t)\dots(\lambda_n - t), \quad \text{for some } c \in \mathbb{C}.$$

Let us also consider part (i) and part (iii) of Lemma 3.

Since we already know that the coefficient of t^n, $b_n = (-1)^n$, we have that $c = 1$.

(i) If we compare coefficients of $t^{(n-1)}$, we have that:

$$(-1)^{n-1}\sum_{i=1}^{n}\lambda_i = (-1)^{(n-1)}tr(A) = b_{(n-1)},$$

and so if we multiply by $(-1)^{n-1}$, we get $\sum_{i=1}^{n}\lambda_i = tr(A) = (-1)^{(n-1)}b_{(n-1)}$.

(ii) If we compare coefficients of t^0, that is the constant terms, we have that:

$$\prod_{i=1}^{n}\lambda_i = b_0 = \det(A). \qquad \blacksquare$$

Corollary 1

Let $A \in M_{n\times n}(\mathbb{C})$.

A is invertible **iff** $\lambda = 0$ is NOT an eigenvalue of A.

Proof: for $A \in M_{n\times n}(\mathbb{C})$.

A is invertible **iff** $\det(A) \neq 0$.
Lemma 4 gives that $\det(A) = \prod_{i=1}^{n}\lambda_i$.
So A is invertible **iff** $\lambda_i \neq 0$ for each eigenvalue $\lambda_i,\ i = 1, \cdots, n$. $\blacksquare$

Lemma 5

Let $A \in M_{n\times n}(\mathbb{F})$.

A is invertible **iff** $\lambda = 0$ is NOT an eigenvalue of A.

Proof: for $A \in M_{n\times n}(\mathbb{F})$, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.

A is invertible **iff** $\det(A) \neq 0$

iff $\det(A - 0\, I_n) \neq 0$ **iff** 0 is not a root of the characteristic polynomial
iff 0 is not an eigenvalue of the matrix A. ■

Examples 13

(i) Example 2 in Topic 16A revisited.

For the matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, we have $\Delta_A(t) = t^2 - 2t - 3$, and $\lambda_1 = 3$, $\lambda_2 = -1$.

We notice that $\Delta_A(t)$ is a quadratic polynomial and the coefficient of t^2 is $(-1)^2 = 1$.

We also have that:

$$\det(A) = -3 = b_0 = \lambda_1 \lambda_2$$

and

$$tr(A) = 2 = \lambda_1 + \lambda_2 = (-1)^1\, b_1 = (-1)(-2).$$

(ii) Example 4 Topic in 16A revisited.

For the matrix $C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, we have $\Delta_C(t) = t^2 + 1$, and $\lambda_1 = i$, $\lambda_2 = -i$.

We notice that $\Delta_C(t)$ is a quadratic polynomial and the coefficient of t^2 is $(-1)^2 = 1$.

We also have that:

$$\det(C) = 1 = b_0 = \lambda_1 \lambda_2$$

and

$$tr(C) = 0 = \lambda_1 + \lambda_2 = (-1)^1\, b_1 = (-1)(0).$$

(iii) Example 5 in Topic 16A revisited.

For the matrix $D = \begin{pmatrix} 3i & -4 \\ 2 & i \end{pmatrix}$, we have that $\Delta_D(t) = t^2 - 4it + 5$, and $\lambda_1 = 5i$, $\lambda_2 = -i$.

We notice that $\Delta_D(t)$ is a quadratic polynomial and the coefficient of t^2 is $(-1)^2 = 1$.

We also have that:

$$\det(D) = 5 = b_0 = \lambda_1 \lambda_2$$

and

$$tr(D) = 4i = \lambda_1 + \lambda_2 = (-1)^1\, b_1 = (-1)(-4i).$$

(iv) Example 6 in Topic $16A$ revisited.

For the matrix $F = \begin{pmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{pmatrix}$, we have that $\Delta_F(t) = -t^3 + 6t^2 + 40t$,

and $\lambda_1 = 10$, $\lambda_2 = 0$, and $\lambda_3 = -4$.

We notice that $\Delta_F(t)$ is a cubic polynomial and the coefficient of t^3 is $(-1)^3 = -1$.

We also have that:

$$\det(F) = 0 = b_0 = \lambda_1\lambda_2\lambda_3$$

and

$$tr(F) = 6 = \lambda_1 + \lambda_2 + \lambda_3 = (-1)^2\, b_2 = (1)(6).$$

(v) Example 8 in Topic 16A revisited.

We have the matrix $Z = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$, and we have that:

$$\Delta_Z(t) = -t^3 + 9t^2 - 24t + 20 = -(t-5)(t-2)^2.$$

We have to be extra careful in this example, since $n = 3$. If we wish to use Lemma 4, then we need to write down the 3 eigenvalues.

These are the three roots of the cubic polynomial and we label them: $\lambda_1 = 5$, $\lambda_2 = 2$, $\lambda_3 = 2$.

We notice that $\Delta_Z(t)$ is a cubic polynomial and the coefficient of t^3 is $(-1)^3 = -1$.

We also have that:

$$\det(Z) = 20 = b_0 = \lambda_1\lambda_2\lambda_3$$

and

$$tr(Z) = 9 = \lambda_1 + \lambda_2 + \lambda_3 = (-1)^2\, b_2 = (1)(9).$$

We conclude this lecture with a result about similar matrices.

Lemma 6

Let $A, B \in M_{n\times n}(\mathbb{F})$.

If A and B are similar, then they have the same characteristic polynomial and the same eigenvalues.

Proof

Since A and B are similar, there exists and invertible matrix Q such that $B = Q^{-1}AQ$. We then have, using Corollaries 8 and 9 in Topic 15B, that:

$$\Delta_B(t) = \det(B - tI) = \det(B - tQ^{-1}IQ)$$

$$= \det(Q^{-1}AQ - tQ^{-1}IQ) = \det(Q^{-1}(A - tI)Q)$$

$$= \det(Q^{-1})\det(A - tI)\det(Q) = \det(Q^{-1})\det(Q)\det(A - tI)$$

$$= \det(A - tI) = \Delta_A(t).$$

We can thus conclude that A and B have the same characteristic polynomial.

Since the eigenvalues of a matrix are just the roots of the characteristic polynomial, we also conclude that A and B have the same eigenvalues. ■

We can apply Lemmas 4 or 6 to conclude that similar matrices have the same trace and the same determinant, and thus we have proved Lemma 1 again.

Topic 17A

Subspaces and Bases

Throughout this course we have been dealing with subsets of the universal sets, $\mathbb{R}^n$ and $\mathbb{C}^n$, for example, solutions sets of systems of equations, lines and planes in $\mathbb{R}^3$, and ranges of linear transformation.

Moving forwards, we will be particularly interested in subsets which have special structure, structure that is appropriate to linear algebra.

The next definition focuses on the properties of that structure.

Definition 1: Subspace

A subset V of $\mathbb{F}^n$ is called a **subspace** of $\mathbb{F}^n$ to mean that:

(i) $\mathbf{0} \in V$

(ii) $\forall \mathbf{x}, \mathbf{y} \in V$, $\mathbf{x} + \mathbf{y} \in V$ (referred to as **closure under addition**).

(iii) $\forall \mathbf{x} \in V$, and $\forall c \in \mathbb{F}$, $c\,\mathbf{x} \in V$ (referred to as **closure under scalar multiplication**).

Comments

The zero vector plays an important role in this course: it is natural that we want it to be in our special subset. This restriction means that the empty set cannot be a subspace.

In addition, if you have a non-empty subset which is closed under scalar multiplication, then the zero vector must be included (since $0(\mathbf{v}) = \mathbf{0}$ for any $\mathbf{v} \in V$).

It is often convenient to know that, when you are "working" in a subset then all the objects that you produce (for instance, by our usual operations of addition (subtraction) and scalar multiplication) will also be in that very same subset. This is why we want to have closure under these operations. To some extent, you may think of the restrictions, that we insist that a subset satisfies in order for it to become a subspace, as linear restrictions. Thus a non-empty subset is a subspace if it is a linear subset.

Lemma 1: Checking for a subspace

Let V be a subset of $\mathbb{F}^n$. Then V is a subspace of $\mathbb{F}^n$ **iff**

(i) V is non-empty

and

(ii) $\forall\, \mathbf{x}, \mathbf{y} \in V$ and $\forall c \in \mathbb{F},\ c\,\mathbf{x} + \mathbf{y} \in V$.

Proof

Let V be a subspace of $\mathbb{F}^n$. Then $\mathbf{0} \in V$, so V is non-empty.

Also $\forall \mathbf{x} \in V$ and $\forall c \in \mathbb{F},\ c\,\mathbf{x} \in V$: let $\mathbf{z} = c\,\mathbf{x}$ if you wish.
Since $\forall\, \mathbf{z}, \mathbf{y} \in V$, we have that $\mathbf{z} + \mathbf{y} \in V$, and it follows that $c\,\mathbf{x} + \mathbf{y} \in V$.

On the other hand, let V be non-empty subset of $\mathbb{F}^n$ and suppose that $\forall\, \mathbf{x}, \mathbf{y} \in V$ and $\forall c \in \mathbb{F},\ c\,\mathbf{x} + \mathbf{y} \in V$.
Then $\exists \mathbf{x} \in V$. Let $c = -1$ and $\mathbf{x} = \mathbf{y}$. Then $(-1)\mathbf{x} + \mathbf{x} \in V$, that is, $\mathbf{0} \in V$.

Also, with $c = 1$, we have that $\forall\, \mathbf{x}, \mathbf{y} \in V$, $1\mathbf{x} + \mathbf{y} \in V$, that is, $\mathbf{x} + \mathbf{y} \in V$,
and, setting $\mathbf{y} = \mathbf{0}$, $\forall\ \mathbf{x} \in V$, and $\forall\ c \in \mathbb{F}$, $c\mathbf{x} + \mathbf{0} \in V$, that is $c\,\mathbf{x} \in V$.

Thus V is a subspace of $\mathbb{F}^n$. $\blacksquare$

Note: this lemma should remind you of Lemma 2 in Topic 13A, which gave an alternate characterization of a linear transformation, and basically allowed us to check linearity under addition and under scalar multiplication "all in one go".

As a consequence of Lemma 2 in Topic 13A, it is relatively easy to check for a subspace.

a) Is it non-empty, and/or containing the zero vector?

b) Does the subset look linear?

If your subset fails either of these tests, then to prove that it is not a subspace,
a counter-example may be needed. If your subset passes both of these tests then it probably is a subspace, so prove it.

Example 1

We have already met many examples of subspaces, and we list many of them here.

(a) $\mathbb{F}^n$ - the largest subspace - the whole set itself.

(b) $\{\mathbf{0}\}$ - the smallest subspace.

$\mathbb{F}^n$ and $\{\mathbf{0}\}$ are collectively known as the **trivial subspaces** of $\mathbb{F}^n$.

(c) Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subset \mathbb{F}^n,\ p \geq 1$, then $Span(\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\})$ is a subspace of $\mathbb{F}^n$.

Proof

Since $\mathbf{v}_1 \in \mathbb{F}^n$, then $(1)\mathbf{v}_1 \in Span(\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\})$, so $Span(\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\})$ is non-empty.

Now suppose that $\mathbf{x}$, $\mathbf{y} \in Span(\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\})$ and $c \in \mathbb{F}$, then there exist scalars $a_1, a_2, \ldots, a_p \in \mathbb{F}$ such that $\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p$, and there exist scalars $b_1, b_2, \ldots, b_p \in \mathbb{F}$ such that $\mathbf{y} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p$.
We then have:

$$c\,\mathbf{x} + \mathbf{y} = c(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_p\mathbf{v}_p) + b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_p\mathbf{v}_p$$

$$= (ca_1 + b_1)\mathbf{v}_1 + (ca_2 + b_2)\mathbf{v}_2 + \cdots + (ca_p + b_p)\mathbf{v}_p,$$

Since the right-hand side of the above equation is an element of $Span(\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\})$, we thus conclude that $c\,\mathbf{x} + \mathbf{y} \in Span(\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\})$.
Therefore, $Span(\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}\})$ is a subspace of $\mathbb{F}^n$.

Many other subsets that we have encountered in our linear algebra journey can be automatically identified as being subspaces of $\mathbb{F}^n$ due to the feature that they may be expressed as $Span(U)$, for some subset U of $\mathbb{F}^n$.

(d) Let $A \in M_{m\times n}(\mathbb{F})$, then $\mathrm{Col}(A)$ is a subspace of $\mathbb{F}^m$.

Proof

$\mathrm{Col}(A) = Span(\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\})$, so this is a special case of (c) above.

(e) Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation, then the range of T, $R(T)$, is a subspace of $\mathbb{F}^m$.

Proof

$R(T) = \mathrm{Col}([T]_S)$, so this is a special case of (d).

(f) Let $A \in M_{m\times n}(\mathbb{F})$, then the solution set to $A\mathbf{x} = \mathbf{0}$, is a subspace of $\mathbb{F}^n$.

Proof

Let Z be the solution set of the system $A\mathbf{x} = \mathbf{0}$.
Since $A\mathbf{0} = \mathbf{0}$, thus $\mathbf{0} \in Z$, and thus Z is non-empty.

Now suppose that $\mathbf{x}$,$\mathbf{y} \in Z$, then $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$.
It now follows, due to the linearity of matrix multiplication (see Lemma 1 in Topic 11A), that

$$A(c\,\mathbf{x} + \mathbf{y}) = c\,A\mathbf{x} + A\mathbf{y} = c\,\mathbf{0} + \mathbf{0} = \mathbf{0},\ \forall c \in \mathbb{F}.$$

Thus $(c\mathbf{x} + \mathbf{y}) \in Z$ and we conclude that Z is a subspace of $\mathbb{F}^n$.

(g) Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation, then $N(T)$, the nullspace of T, is a subspace of $\mathbb{F}^n$.

Proof

$N(T)$ is identical to the solution set of the homogeneous linear systems of equations, $[T]_S\,\mathbf{x} = \mathbf{0}_{\mathbb{F}^m}$, (see Lemma 5 in Topic 13A). Thus $N(T)$ is a subspace of $\mathbb{F}^n$ by (f).

(h) Let $A \in M_{n\times n}(\mathbb{F})$ and let λ be an eigenvalue of A, then E_λ is a subspace of $\mathbb{F}^n$.

Proof

E_λ is equal to the solution set to the homogeneous linear systems of equations, $(A - \lambda I)\mathbf{x} = \mathbf{0}$, and this is a subspace of $\mathbb{F}^n$ by (f).

Example 2

Let $W_1 = \{(x, y, z)^T \in \mathbb{R}^3 : x + 2y + 3z = 4\}$ be a subset of $\mathbb{R}^3$. Is W_1 a subspace of $\mathbb{R}^3$?

Solution

No, W_1 is not a subspace of $\mathbb{R}^3$, since $(0, 0, 0)^T \notin W_1$.

Let $W_2 = \{(x, y, z)^T \in \mathbb{R}^3 : x^2 - z^2 = 0\}$ be a subset of $\mathbb{R}^3$. Is W_2 a subspace of $\mathbb{R}^3$?

Solution

Although $(0, 0, 0)^T \in W_2$, this subset does not look linear. Let us check closure under addition.
Let us consider some vectors avoiding all components which are 0 and 1.
Clearly $(2, 0, 2)^T \in W_2$, as $2^2 - 2^2 = 0$, and $(-2, 0, 2)^T \in W_2$, as $(-2^2) - 2^2 = 0$, however if we add these two vectors together, we get $(2, 0, 2)^T + (-2, 0, 2)^T = (0, 0, 4)^T$, and since $(0^2) - 4^2 = -16 \neq 0$, then $(0, 0, 4)^T \notin W_2$.
Closure under addition is not holding and thus W_2 is not a subspace of $\mathbb{R}^3$.

What are the subspaces of $\mathbb{F}^n$?

Example 3

(a) For $\mathbb{F}$, there are just two subspaces: $\{\mathbf{0}\}$ and $\mathbb{F}$, the two trivial subspaces. Why?

Let us assume there is another subspace: either there is just $\mathbf{0}$ in that subspace, in which case the subspace is $\{\mathbf{0}\}$, or there is at least another non-zero element, call this $\mathbf{x}$. Any other element of $\mathbb{F}$, $\mathbf{y}$ say, can be written as $\mathbf{y} = m\,\mathbf{x}$, for some $m \in \mathbb{F}$. Since a subspace is closed under scalar multiplication, if $\mathbf{x}$ is an element of your subspace, then $\mathbf{y}$ is also an element of your subspace too. We conclude that your subspace is all of $\mathbb{F}$.

(b) For $\mathbb{F}^2$, there are: $\{\mathbf{0}\}$ and $\mathbb{F}^2$, the two trivial subspaces. Are there any more?

Let $\mathbf{x} \in \mathbb{F}^2, \mathbf{x} \neq \mathbf{0}$. Then $Span(\{\mathbf{x}\})$ is a straight line through the origin. This is a subspace of $\mathbb{F}^2$ by Example 1 (c).

$Span(\{\mathbf{x}\})$ is also the smallest subspace of $\mathbb{F}^2$ that contains the vector $\mathbf{x}$, in the sense that if you have a subspace of $\mathbb{F}^2$ that contains $\mathbf{x}$, then it contains $Span(\{\mathbf{x}\})$. This follows by closure of a subspace under scalar multiplication. Are there any others?

Now suppose that we have $\mathbf{x}, \mathbf{y} \in \mathbb{F}^2$, then $Span(\{\mathbf{x}, \mathbf{y}\})$ is the smallest subspace of $\mathbb{F}^2$ which contains both $\mathbf{x}$ and $\mathbf{y}$. That means that if you have a subspace of $\mathbb{F}^2$ that contains both $\mathbf{x}$ and $\mathbf{y}$, then it contains $Span(\{\mathbf{x}, \mathbf{y}\})$.

If, in addition, $\mathbf{x} \neq \mathbf{0}$, and $\mathbf{y} \in Span(\{\mathbf{x}\})$, then $Span(\{\mathbf{x}, \mathbf{y}\}) = Span(\{\mathbf{x}\})$, so that in this case, the addition of the vector $\mathbf{y}$ did not gain us anything.

Now suppose we have $\mathbf{x} \neq \mathbf{0}$, and $\mathbf{y} \notin Span(\{\mathbf{x}\})$, then $Span(\{\mathbf{x}, \mathbf{y}\}) = \mathbb{F}^2$. We justify this statement. We claim these two sets are equal. We already know that $Span(\{\mathbf{x}, \mathbf{y}\}) \subset \mathbb{F}^2$, so we need only show that $\mathbb{F}^2 \subset Span(\{\mathbf{x}, \mathbf{y}\})$. Suppose that we write $\mathbf{x} = (x_1, x_2)^T$ and $\mathbf{y} = (y_1, y_2)^T$, with $(x_1, x_2)^T \neq (0, 0)^T$ and $\mathbf{y} \neq m\mathbf{x}$, for any $m \in \mathbb{F}$. We show that if $\mathbf{z} = (z_1, z_2)^T \in \mathbb{F}^2$, then $\mathbf{z} \in Span(\{\mathbf{x}, \mathbf{y}\})$. That is, we show that we can solve $\mathbf{z} = a\mathbf{x} + b\mathbf{y}$ for some $a, b \in \mathbb{F}$. The augmented matrix is shown below.

$$\left(\begin{array}{cc|c} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{array}\right).$$

If $x_1 = 0$, then $y_1 \neq 0$ since $\mathbf{y} \notin Span(\{\mathbf{x}\})$. Thus, the coefficient matrix has two pivots, and thus we have a (unique) solution.

If $x_1 \neq 0$, then row reducing the augmented matrix (without showing here all the steps), will yield:

$$\left(\begin{array}{cc|c} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & \dfrac{y_1}{x_1} & \dfrac{z_1}{x_1} \\ 0 & \dfrac{y_2x_1 - x_2y_1}{x_1} & \dfrac{z_2x_1 - x_2z_1}{x_1} \end{array}\right).$$

In this case, $y_2x_1 - x_2y_1 = 0$ iff $(y_1, y_2)^T = \dfrac{y_1}{x_1}(x_1, x_2)^T$ which we exclude, since $\mathbf{y} \notin Span(\{\mathbf{x}\})$. And so the coefficient matrix has two pivots, and thus we have a (unique) solution.

Thus the only subspaces of $\mathbb{F}^2$ are the two trivial ones, together with lines through the origin. The latter can be written as $Span(\{\mathbf{x}\})$, with $\mathbf{x} \neq \mathbf{0}$.

(c) For $\mathbb{F}^3$, there are: $\{\mathbf{0}\}$ and $\mathbb{F}^3$ the two trivial subspaces. Are there any more?

Let $\mathbf{x} \in \mathbb{F}^3, \mathbf{x} \neq \mathbf{0}$. Then $Span(\{\mathbf{x}\})$ is a straight line through the origin. This is a subspace of $\mathbb{F}^3$ by Example 1 (c), and is the smallest subspace that contains the vector $\mathbf{x}$.

Let $\mathbf{x}, \mathbf{y} \in \mathbb{F}^3$. Then $Span(\{\mathbf{x}, \mathbf{y}\})$ is the smallest subspace of $\mathbb{F}^3$ which contains both $\mathbf{x}$ and $\mathbf{y}$.

Suppose now that, in addition, $\mathbf{x} \neq \mathbf{0}$, then either $\mathbf{y} \in Span(\{\mathbf{x}\})$, or $\mathbf{y} \notin Span(\{\mathbf{x}\})$.
If $\mathbf{y} \in Span(\{\mathbf{x}\})$, then $\mathbf{y}$ contributes nothing new and $Span(\{\mathbf{x}, \mathbf{y}\}) = Span(\{\mathbf{x}\})$, a straight line through the origin.
If $\mathbf{y} \notin Span(\{\mathbf{x}\})$, then $Span(\{\mathbf{x}, \mathbf{y}\})$ is a plane through the origin in $\mathbb{F}^3$.
If we now were to add another vector $\mathbf{z} \in \mathbb{F}^3$, then, not surprisingly, the smallest subspace which contains $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ is $Span(\{\mathbf{x}, \mathbf{y}, \mathbf{z}\})$. Now either $\mathbf{z} \in Span(\{\mathbf{x}, \mathbf{y}\})$, or $\mathbf{z} \notin Span(\{\mathbf{x}, \mathbf{y}\})$.
In the former case, then $\mathbf{z}$ contributes nothing new and $Span(\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = Span(\{\mathbf{x}, \mathbf{y}\})$, and in the latter case, we **can** show that $Span(\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = \mathbb{F}^3$.

Thus the only subspaces of $\mathbb{F}^3$ are the two trivial ones, together with lines through the origin, which can be written as $Span(\{\mathbf{x}\})$, with $\mathbf{x} \neq \mathbf{0}$, and planes through the origin, which can be written as $Span(\{\mathbf{x}, \mathbf{y}\})$, with $\mathbf{x} \neq \mathbf{0}$, and $\mathbf{y} \notin Span(\{\mathbf{x}\})$.

We will now concentrate on the issue of how we *describe a subspace in an efficient way.* It would be a good idea to review the introductory lecture on sets, where, among other topics, we emphasized the important idea of there being many different ways to write down the exact same set. The method of writing down all the elements of the subset (subspace) is clearly just not a possibility as there are too many vectors to write them all down. Even if we could write them all down, we would have a very large set of vectors to carry around with us. We have already seen that the technique of selecting a small set of vectors, S, and using the method of taking their span, i.e. forming $V = Span(S)$, is a very simple and powerful way of producing a large set V from this small set S. We will make use of this process, and we have already had some experience with it. However, another issue we will have to resolve is what vectors should we be putting in this set S, which we will use to build V. Once again, we will insist that we that we avoid redundancy.

Example 4

Suppose we tell you that we have a plane, P through the origin in $\mathbb{R}^3$.

We are going to describe this plane as $\mathit{Span}(S)$, for some subset S of $\mathbb{R}^3$.

You already are aware that we cannot let $S = \{\mathbf{v}\}$, for some non-zero vector $\mathbf{v} \in \mathbb{R}^3$ and thus describe P as $\mathit{Span}(\{\mathbf{v}\})$. The span of a single non-zero vector is a line, not a plane. We have not given you enough to work with.

So, now suppose that we tell you that we have a subset, S, of $\mathbb{R}^3$,

$$S = \left\{ \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 13 \\ -8 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

In this case, $\mathit{Span}(S)$ could be a plane. In fact, we know from our earlier discussion, that since there are, at least two non-zero, non-parallel vectors in S, then it must be at least a plane. However, $\mathit{Span}(S)$ could even be the entire $\mathbb{R}^3$.

If $\mathit{Span}(S)$ is a plane, then we know that we can be much more efficient in describing this plane. It is possible to provide a smaller set, by which I mean a set with fewer vectors in it, let us call it S_2, such that $\mathit{Span}(S_2) = P$. Even further, it is possible to choose S_2 to be a subset of S. How do we know which vectors to select, and which vectors are redundant in S? This leads us to the important idea of linear dependence, and linear independence.

Informally we say that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, are linearly dependent to mean that at least one of them, $\mathbf{v}_k$ say, can be written as a linear combination of some of the other vectors. The vector $\mathbf{v}_k$ depends linearly on the other vectors.

For example, the set S above is linearly dependent because,

$$\begin{pmatrix} -2 \\ 13 \\ -8 \end{pmatrix} = 2 \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} - 3 \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}.$$

There is a good reason that we do not use this definition as our formal definition, and that is, we have singled out the vector $(-2, 13, -8)^T$. There is no clear reason why we should do this: for example, we could also write

$$\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -2 \\ 13 \\ -8 \end{pmatrix} + \frac{3}{2} \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}$$

or again

$$\begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} = \frac{2}{3} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} -2 \\ 13 \\ -8 \end{pmatrix}.$$

It seems more reasonable and balanced if we put everything on one side of the equation and write the equation in the following way:

$$\begin{pmatrix} -2 \\ 13 \\ -8 \end{pmatrix} - 2 \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} + 3 \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The issue of redundancy in the set S is identified by the feature that we can combine some of the vectors in S, with a linear combination, and produce the zero vector. When we try to convert this idea into a definition, we will have to be a little careful. We can always make the zero vector by taking the **trivial linear combination**, i.e.

$$0 \begin{pmatrix} -2 \\ 13 \\ -8 \end{pmatrix} + 0 \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} + 0 \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We have to eliminate this simple way of making zero from our formal definition.

Definition 2: Linear dependence

We say that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are **linearly dependent** to mean that there exists scalars $c_1, c_2, \ldots, c_p$, not all zero, such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$.

That is, the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are linearly dependent if it is possible to find a non-trivial linear combination of them which produces the zero vector.

If $U = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, then we say that the set is a **linearly dependent set** (or simply the set is linearly dependent) to mean that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are linearly dependent.

Definition 3: Linear independence

We say that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are linearly independent to mean that there do not exist scalars $c_1, c_2, \ldots, c_p$, not all zero, such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$.

That is, the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are linearly independent if the only possible linear combination of them which produces the zero vector, is the trivial linear combination, i.e. $c_1 = c_2 = \cdots = c_p = 0$.

If $U = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, then we say that the set is a **linearly independent set** (or simply the set is linearly independent) to mean that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are linearly independent.

When we are given a set that is linearly dependent, we immediately think of that set as being deficient in some way. We have been supplied with more vectors than we really need, and we could remove at least one vector, $\mathbf{v}_k$ say, from the set, in order to "lighten my load." If we do remove the vector $\mathbf{v}_k$, then we have a smaller set of vectors, but we can construct $\mathbf{v}_k$ from the remaining vectors, through a linear combination. Note that you will usually have some choice about which vector you remove.

Let us return to Example 4 of the plane P through the origin in $\mathbb{R}^3$.

$$P = Span(S) = Span\left\{ \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 13 \\ -8 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

We know that S is linearly dependent. We choose to remove the vector $(-2, 13, -8)^T$ from S to produce the set S_1. Then

$$P = Span(S_1) = Span\left\{ \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

However S_1 is linearly dependent, for example:

$$\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} - 2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We choose to remove the vector $(2, 2, 2)^T$ from S_1 to produce the set S_2.

$$P = Span(S_2) = Span\left\{ \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 5 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

However S_2 is linearly dependent, for example:

$$2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} - \begin{pmatrix} 0 \\ 5 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We choose to remove the vector $(0, 5, -2)^T$ from S_2 to produce the set S_3.

$$P = Span(S_3) = Span\left\{ \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

The set S_3 is linearly independent, we cannot reduce the spanning set any further. The set S_3 is an optimal way of producing the plane P.

Suppose that we have a large set V, this could be a subspace of $\mathbb{F}^n$, for example.
It is often possible to build the large set V from a much smaller set, U say, by the process of spanning the set U. We can check that the smaller set, U, does not contain redundant vectors, by making sure that U is linearly independent. If the set U satisfies both of these properties, then U provides an optimal way of constructing the set V through the process of linear combinations. We consider the set U as being a basic set of building blocks for the construction of the set V.

Definition 4: Basis

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ be a subset of vectors contained in the subspace V of $\mathbb{F}^n$.
We say that B is a **basis** for V to mean that B is a linearly independent set of vectors which spans V.

Thus B is a basis for V means that:

(i) $B \subseteq V$

(ii) $Span(B) = V$

(iii) B is linearly independent.

TOPIC 17B
Linear Independence

We begin with some results about testing a set S to see whether the set is linearly dependent, or linearly independent.

First of all, just in case you were wondering, the empty set $\emptyset$ is linearly independent, since you cannot build the zero vector, $\mathbf{0}$, from the empty set and linear combinations.

Lemma 2

If $\mathbf{0} \in S \subseteq \mathbb{F}^n$, then S is linearly dependent.

Proof

Since $1 \neq 0$, and $1(\mathbf{0}) = \mathbf{0}$, we can thus take a non-trivial linear combination of some vectors (in this case, only one vector) in S to produce the zero vector. ■

Lemma 3

If $S \subseteq \mathbb{F}^n$, and S contains a single vector, $\mathbf{x}$, then S is linearly dependent **iff** $\mathbf{x} = \mathbf{0}$.

Proof

We know from Lemma 2 that if $\mathbf{x} = \mathbf{0}$, then S is linearly dependent.

In addition, we know from properties of zero (Lemma 4 in Topic 1) that:

$$\text{If } c\,\mathbf{x} = \mathbf{0}, \text{ then either } \mathbf{x} = \mathbf{0} \text{ or } c = 0.$$

Thus, if $\mathbf{x} \neq \mathbf{0}$, and $c\,\mathbf{x} = \mathbf{0}$, then c must be zero, that is, only the trivial combination of $\mathbf{x} \neq \mathbf{0}$ will produce the zero vector. Thus S is linearly independent. ■

Thus, if we have a set, S, with only one vector in it, it is easy to check for linear independence. The set, S, is linearly dependent iff $S = \{\mathbf{0}\}$.

What about a set with two vectors in it? Well, first of all look and see if one of them is the zero vector. If it is, then the set is linearly dependent.

Lemma 4

If $S \subseteq \mathbb{F}^n$, and $S = \{\mathbf{x}, \mathbf{y}\}$, then S is linearly dependent **iff** one of the vectors is a multiple of the other.

Proof

If one of the vectors is a multiple of the other, then we can write w.l.o.g., that $\mathbf{y} = m\mathbf{x}$. We then have $\mathbf{y} - m\,\mathbf{x} = \mathbf{0}$, and since the coefficient of $\mathbf{y}$ is 1, we have demonstrated linear dependence.

On the other hand, if $S = \{\mathbf{x}, \mathbf{y}\}$ is linearly dependent, then there exist $a, b \in \mathbb{F}$, not both zero, such that $a\,\mathbf{y} + b\,\mathbf{x} = \mathbf{0}$. If we assume, w.l.o.g., that $a \neq 0$, we then have

$$\mathbf{y} + \frac{b}{a}\,\mathbf{x} = \mathbf{0}, \text{ that is , } \mathbf{y} = -\frac{b}{a}\,\mathbf{x}.$$

In other words, one of the vectors can be written as a multiple of the other. ■

Thus if you have a set consisting of only two vectors then it is relatively easy to check it for linear dependence. First see if one of them is the zero vector. If it is not, then see if one is a multiple of the other. This may not be too obvious when the field is $\mathbb{C}$.

Note that in a theoretical setting, if we have a set $S = \{\mathbf{x}, \mathbf{y}\}$, and we are told that it is linearly dependent, then we know that one of the vectors can be written in terms of the other. We do not know which one though.

For example, if $\mathbf{x} = \mathbf{0}$ and $\mathbf{y} \neq \mathbf{0}$, then $\mathbf{y} \neq m\,\mathbf{x}$, for any scalar m, but you do not know this! In such a situation, you should just state that w.l.o.g. we can assume that $\mathbf{y} = m\,\mathbf{x}$, for some scalar m **or** you can assume, w.l.o.g., that $\mathbf{x} = m\,\mathbf{y}$, for some scalar m.

Example 5

Is $S = \left\{ \begin{pmatrix} 2 \\ -4 \end{pmatrix}, \begin{pmatrix} 6 \\ -12 \end{pmatrix} \right\}$ linearly dependent?

Solution

Yes of course, it is: we see immediately that the second vector is three times the first one.

Example 6

Is $S = \left\{ \begin{pmatrix} 2 \\ -4 \end{pmatrix}, \begin{pmatrix} 6 \\ -14 \end{pmatrix} \right\}$ linearly dependent?

Solution

No, of course, it is not: we see immediately that the second vector is not a multiple of the first one. Looking at the first components, 6 is three times 2, however, looking at the second components, -14 is not 3 times -4.

Example 7

Is $S = \{\mathbf{z}_1, \mathbf{z}_2\}$, with $\mathbf{z}_1 = \begin{pmatrix} 2+3i \\ 4+i \end{pmatrix}$ and $\mathbf{z}_2 = \begin{pmatrix} 12+5i \\ 14-5i \end{pmatrix}$, linearly dependent?

Solution

This is not obvious at all, because we are working over $\mathbb{C}$.

We can either try to detect a multiplicative factor or just use the definition.

We will use the definition. Consider $a\,\mathbf{z}_1 + b\,\mathbf{z}_2 = \mathbf{0}$.

Let us row reduce the corresponding coefficient matrix:

$$\begin{pmatrix} 2+3i & 12+5i \\ 4+i & 14-5i \end{pmatrix} \xrightarrow{R_1 \to \frac{1}{2+3i} R_1 = \frac{2-3i}{13} R1}$$

$$\begin{pmatrix} 1 & 3-2i \\ 4+i & 14-5i \end{pmatrix} \xrightarrow{R_2 \to R_2 - (4+i) R_1} \begin{pmatrix} 1 & 3-2i \\ 0 & 0 \end{pmatrix}.$$

We deduce from this reduced matrix that there are (infinitely many) non-trivial solutions to the system of equations, for example: $a = -3+2i$ and $b = 1$.
Thus the set S is linearly dependent.

How do we deal with a set that has more than two vectors?

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subseteq \mathbb{F}^n$, we consider the following questions about S
(a) Is $\mathbf{0} \in S$?
(b) Is one vector in S a multiple of another vector in S?
(c) Do you observe that one vector in S can be written as a linear combination of some of the other vectors in S?

If you answer yes to any of these questions, then the set S is linearly dependent.

If you do not answer yes to any of these question, then you do not know whether the set S is linearly dependent or linearly independent, and you need to use the definition.

Thus we must examine the expression $c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p = \mathbf{0}$.

This is a homogeneous linear system of n equations in p unknowns, which we are most capable in solving. Since it is a homogeneous system, we know that one solution is the trivial solution, and we wish to know whether or not there are any other solutions.

If there are non-trivial solutions, then the set S is linearly dependent.

Usually we do not actually need the solutions, unless we are explicitly asked to obtain them, we just need to know whether or not they exist.

The answer to this question is determined entirely by the rank of the coefficient matrix of this system.

Lemma 5

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subseteq \mathbb{F}^n$, and let $A = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$ be an $(n \times p)$ matrix having $rank(A) = r$ and pivot columns labeled $q_1, q_2, \ldots, q_r$.

Let $U = \{\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}\}$. Then

(a) S is linearly independent **iff** $r = p$.

(b) U is linearly independent.

(c) A subset, V, of S that contains U and any other vector from S, is linearly dependent.

(d) $Span(U) = Span(S)$.

Proof

The matrix A is the coefficient matrix for the homogeneous system of linear equations

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p = \mathbf{0}.$$

There will no parameters in the solution set **iff** there is a pivot in each of the columns of A, that is, **iff** $r = p$. So S is linearly independent **iff** $r = p$. Part (a) is thus proved.

Now suppose that $r < p$, so that S is linearly dependent. Let us select the vectors corresponding to the pivot columns, $\mathbf{v}_{q_1}, \mathbf{v}_{q2}, \ldots, \mathbf{v}_{q_r}$, and let us examine them for linear dependence. That is, we must consider the homogeneous system of linear equations:

$$d_1\,\mathbf{v}_{q_1} + d_2\,\mathbf{v}_{q_2} + \cdots + d_r\,\mathbf{v}_{q_r} = \mathbf{0}.$$

This system will have coefficient matrix equal to $(\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r})$. We already know that this matrix has a rank of r, because it is part of the matrix A that we have already row reduced. We conclude that the only solution of this equation is the trivial solution, and so, the set U is linearly independent. Part (b) is thus proved.

In addition, suppose that column k is **not** a pivot column, and we add the corresponding vector, $\mathbf{v}_k$, to the set U, and check this new set for linear dependence. We consider the system of equations :

$$d_1 \mathbf{v}_{q_1} + d_2 \mathbf{v}_{q_2} + \cdots + d_r \mathbf{v}_{q_r} + \alpha \mathbf{v}_k = \mathbf{0}.$$

This system will have coefficient matrix $(\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}, \mathbf{v}_k)$, which is the r pivot columns of A, together with the k^{th} column of A. Thus this matrix has a rank of r, with the r pivots being in the first r columns, and so we can choose α to be the parameter. That is, we can write the vector $\mathbf{v}_k$ as a linear combination of the vectors $\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}$. Thus the addition of the vector $\mathbf{v}_k$ to the set U produces a new set which is linearly dependent.
Part (c) is now proved.

Assume that $\mathbf{v}_k = a_1\mathbf{v}_{q_1} + a_2\mathbf{v}_{q_2} + \cdots + a_r\mathbf{v}_{q_r}$. If $\mathbf{w} \in Span(\{\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}, \mathbf{v}_k\})$, then

$$\begin{aligned}\mathbf{w} &= b_1 \mathbf{v}_{q_1} + b_2 \mathbf{v}_{q_2} + \cdots + b_r \mathbf{v}_{q_r} + c\mathbf{v}_k, \quad \text{for some scalars } b_1, b_2, \cdots, b_r \text{ and } c\\ &= b_1 \mathbf{v}_{q_1} + b_2 \mathbf{v}_{q_2} + \cdots + b_r \mathbf{v}_{q_r} + c\,(a_1\mathbf{v}_{q_1} + a_2\mathbf{v}_{q_2} + \cdots + a_r\mathbf{v}_{q_r})\\ &= (b_1 + c\,a_1)\,\mathbf{v}_{q_1} + (b_2 + c\,a_2)\,\mathbf{v}_{q_2} + \cdots + (b_r + c\,a_r)\,\mathbf{v}_{q_r}.\end{aligned}$$

Thus, $\mathbf{w} \in Span(\{\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}\})$ and

$$Span(\{\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}\}) = Span(\{\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}, \mathbf{v}_k\}).$$

Repeating this for all the non-pivot columns yields the result that $Span(U) = Span(S)$.
Part (d) is now proved. ■

Lemma 5 makes the identification of linearly independent/dependent subsets relatively straightforward. In addition, we have now a method for reducing them to linearly independent sets which have the same span as the original set.

Corollary 1

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subseteq \mathbb{F}^n$. If $n < p$, then S is linearly dependent.

Proof

The coefficient matrix for the system

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p = \mathbf{0}$$

will be a matrix with n rows and p columns. Suppose that this matrix has rank of r. We know that both $r \le p$ and $r \le n$. Thus, if $n < p$, then $r < p$, and the result follows from Lemma 5 part (c): that is, if you have more than n vectors in a subset of $\mathbb{F}^n$, then they are linearly dependent. ■

Example 8

Consider the following vectors in $\mathbb{C}^6$:

$\mathbf{v}_1 = (1, -2, 3, -4, 5, -6)^T,\ \mathbf{v}_2 = (3, 4, 5, 6, 7, 8)^T,\ \mathbf{v}_3 = (-5, -10, -7, -16, -9, -22)^T.$

$\mathbf{v}_4 = (-6, -4, -2, 0, 2, 4)^T,\ \mathbf{v}_5 = (1, 1, 1, 1, 1, 1)^T$, and $\mathbf{v}_6 = (-3, 3, -5, 5, -1, 1)^T$.

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_6\}$ and $V = Span(S)$.

(a) Show that S is linearly dependent.

(b) Find a subset of S which is a basis for V.

(c) Find all the other linearly independent subsets of S.

Solution

We consider the equation $d_1\,\mathbf{v}_1 + d_2\,\mathbf{v}_2 + d_3\,\mathbf{v}_3 + d_4\,\mathbf{v}_4 + d_5\,\mathbf{v}_5 + d_6\,\mathbf{v}_6 = \mathbf{0}$.

The coefficient matrix for this system is :

$$A = \begin{pmatrix} 1 & 3 & -5 & -6 & 1 & -3 \\ -2 & 4 & -10 & -4 & 1 & 3 \\ 3 & 5 & -7 & -2 & 1 & -5 \\ -4 & 6 & -16 & 0 & 1 & 5 \\ 5 & 7 & -9 & 2 & 1 & -1 \\ -6 & 8 & -22 & 4 & 1 & 1 \end{pmatrix}.$$

A row echelon form of this matrix is:

$$\begin{pmatrix} 1 & 3 & -5 & -6 & 1 & -3 \\ 0 & 1 & -2 & \frac{-8}{5} & \frac{3}{10} & \frac{-3}{10} \\ 0 & 0 & 0 & 1 & \frac{-1}{12} & \frac{7}{24} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The reduced row echelon form of the matrix is:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & 1 & \frac{-1}{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We can see that the *rank* of A is 4, and so there will be two parameters in the solution set to $d_1\,\mathbf{v}_1 + d_2\,\mathbf{v}_2 + d_3\,\mathbf{v}_3 + d_4\,\mathbf{v}_4 + d_5\,\mathbf{v}_5 + d_6\,\mathbf{v}_6 = \mathbf{0}$: and thus, S is linearly dependent. Part (a) is now proved.

Lemma 5 tells us that the pivot columns $1, 2, 4$, and 6, are linearly independent and that their span is the same of S. Thus the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$ and $\mathbf{v}_6$ are:

(i) in V,
(ii) linearly independent, and
(iii) $Span(\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4,\ \mathbf{v}_6\}) = V$.

We conclude that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4,\ \mathbf{v}_6\}$ is a basis for V. Part (b) is now proved.

Part (c) is an interesting question which demands some thought, but the entire solution to this part can be read off from $RREF(A)$. We know already that the set S of six vectors is linearly dependent.

If you take any five of these vectors and consider the equation:

$$d_1\,\mathbf{v}_{q_1} + d_2\,\mathbf{v}_{q_2} + \cdots + d_5\,\mathbf{v}_{q_5} = \mathbf{0}, \text{ where } 1 \le q_1 \le q_2 \le \cdots \le q_5 \le 6,$$

then the coefficient matrix of this system has a *rank* that is less than or equal to $rank(A) = 4$. Thus the vectors will be linearly dependent by Lemma 5.

If you take four of these vectors and consider the equation:

$$d_1\mathbf{v}_{q_1} + d_2\mathbf{v}_{q_2} + d_3\mathbf{v}_{q_3} + d_4\mathbf{v}_{q_4} = \mathbf{0}, \text{ where } 1 \le q_1 \le q_2 \le q_3 \le q_4 \le 6,$$

then the coefficient matrix of this system has a *rank* that is less than or equal to $rank(A) = 4$. Thus the 4 vectors could be linearly independent. This will depend on which 4 vectors you actually select. For example, we know that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4,\ \mathbf{v}_6\}$ is linearly independent. However, if you choose the first four vectors, then your coefficient matrix would be

$$B = \begin{pmatrix} 1 & 3 & -5 & -6 \\ -2 & 4 & -10 & -4 \\ 3 & 5 & -7 & -2 \\ -4 & 6 & -16 & 0 \\ 5 & 7 & -9 & 2 \\ -6 & 8 & -22 & 4 \end{pmatrix},$$

a row echelon form of this is

$$\begin{pmatrix} 1 & 3 & -5 & -6 \\ 0 & 1 & -2 & \frac{-8}{5} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

which only has 3 pivots. Thus those four vectors are linearly dependent.

The following choices of 4 vectors, and only these choices will yield 4 linearly independent vectors. In each case, an examination of the coefficient matrix for the corresponding homogeneous system of equations is required in order to see whether or not it has 4 pivots.

$$\begin{gathered}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_6\}, \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5, \mathbf{v}_6\}, \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_6\}, \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_5, \mathbf{v}_6\} \\ \{\mathbf{v}_1, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}, \{\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_6\}, \{\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_5, \mathbf{v}_6\}, \{\mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}.\end{gathered}$$

It is possible to re-order the vectors in S so that, for any single choice of four linearly independent vectors, the columns, corresponding to your choice, are the pivot columns. Thus any one of these sets is a basis for V.

Now suppose that we wanted to choose 3 vectors from S and we want them to be linearly independent. We would have to examine the system:

$$d_1\,\mathbf{v}_{q_1} + d_2\,\mathbf{v}_{q_2} + d_3\,\mathbf{v}_{q_3} = \mathbf{0}, \text{ where } 1 \le q_1 \le q_2 \le q_3 \le 6.$$

And the coefficient matrix for this homogeneous linear system would consist of three columns of the matrix A. We need to arrange for each of these columns to be a pivot column. We could not, for example, choose $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$.

Careful examination of $RREF(A)$ reveals that we can choose $\mathbf{v}_6$ and any other two vectors from the set S. Or we can choose both $\mathbf{v}_4$ and $\mathbf{v}_5$, and either $\mathbf{v}_1$ or $\mathbf{v}_3$. Or, we can choose either of $\mathbf{v}_4$ or $\mathbf{v}_5$ and any two of $\mathbf{v}_1$, $\mathbf{v}_2$ or $\mathbf{v}_3$. Any one of these sets are linearly independent. We get 18 sets in total.

Now suppose that we wanted to choose 2 vectors from S and we want them to be linearly independent. Since none of the vectors is the zero vector, and a quick examination shows us that none is a multiple of the other, we conclude that we can select any two of the vectors in S to get a linearly independent subset of S. We get 15 sets in total.

Finally, any subset of S consisting of one vector will be linearly independent, and the empty set is also linearly independent.

That gives a total of 48 different linearly independent subsets of S.

Example 9

Consider the following vectors in $\mathbb{C}^5$.

$\mathbf{v}_1 = (1+i, -2+i, 3, -4-i, 5i)^T$, $\mathbf{v}_2 = (-1+5i, -7-4i, 6+9i, -5-14i, -15+10i)^T$,

$\mathbf{v}_3 = (3i, 4+2i, 5+2i, 6-4i, 7-2i)^T$, $\mathbf{v}_4 = (-1+2i, -3-6i, 11+9i, 1-18i, -8+10i)^T$,

and $\mathbf{v}_5 = (-6i, -4i, -2i, 0, 2i)^T$.

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_5\}$ and $V = Span(S)$.

(a) Show that S is linearly dependent.

(b) Find a subset of S which is a basis for V.

(c) Find all the other linearly independent subsets of S.

Solution

We consider the equation $d_1\,\mathbf{v}_1 + d_2\,\mathbf{v}_2 + d_3\,\mathbf{v}_3 + d_4\,\mathbf{v}_4 + d_5\,\mathbf{v}_5 = \mathbf{0}$.

The coefficient matrix for this system is

$$A = \begin{pmatrix} 1+i & -1+5i & 3i & -1+2i & -6i \\ -2+i & -7-4i & 4+2i & -3-6i & -4i \\ 3 & 6+9i & 5+2i & 11+9i & -2i \\ -4-i & -5-14i & 6-4i & 1-18i & 0 \\ 5i & -15+10i & 7-2i & -8+10i & 2i \end{pmatrix},$$

and the $RREF(A)$ is

$$\begin{pmatrix} 1 & 2+3i & 0 & 0 & -2-3i \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We can see that the *rank* of A is 3, and so there will be two parameters in the solution set to $d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + d_3\mathbf{v}_3 + d_4\mathbf{v}_4 + d_5\mathbf{v}_5 = \mathbf{0}$, and thus S is linearly dependent. Part (a) is now proved.

Lemma 5 tells us that the pivot columns $1, 3$, and 4 are linearly independent, and the span of the corresponding vectors is the same as $Span(S)$.
Thus the vectors $\mathbf{v}_1, \mathbf{v}_3$ and $\mathbf{v}_4$:

(i) are in V,

(ii) are linearly independent, and
(iii) $Span(\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4\}) = V$.

We conclude that $\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4\}$ is a basis for V, and part (b) is solved.

Part (c) is an interesting question which demands some thought, but the entire solution to this can be read off from $RREF(A)$. We know already that the set S of five vectors is linearly dependent.

If you take any four of these vectors and consider the equation:

$$d_1\,\mathbf{v}_{q_1} + d_2\,\mathbf{v}_{q_2} + \cdots + d_4\,\mathbf{v}_{q_4} = \mathbf{0}, \text{ where } 1 \le q_1 \le q_2 \le \cdots \le q_4 \le 5,$$

then the coefficient matrix of this system has a *rank* that is less than or equal to $rank(A) = 3$. Thus the vectors will be linearly dependent by Lemma 5.

If you take three of these vectors and consider the equation:

$$d_1\mathbf{v}_{q_1} + d_2\mathbf{v}_{q_2} + d_3\mathbf{v}_{q_3} = \mathbf{0}, \text{ where } 1 \le q_1 \le q_2 \le q_3 \le 5,$$

then the coefficient matrix of this system has a *rank* that is less than or equal to $rank(A) = 3$. Thus the 3 vectors could be linearly independent. This will depend on which 3 vectors you actually select. For example, we know that $\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent. However, if you choose the first three vectors, then your coefficient matrix would be

$$B = \begin{pmatrix} 1+i & -1+5i & 3i \\ -2+i & -7-4i & 4+2i \\ 3 & 6+9i & 5+2i \\ -4-i & -5-14i & 6-4i \\ 5i & -15+10i & 7-2i \end{pmatrix},$$

for which the RREF is:

$$\begin{pmatrix} 1 & 2+3i & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which only has 2 pivots. Thus those three vectors are linearly dependent.

The following choices of 3 vectors, and only these choices will yield 3 linearly independent vectors. In each case, an examination of the coefficient matrix for the corresponding homogeneous system of equations is required in order to see whether or not it has 3 pivots.

$$\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4\}, \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_5\}, \{\mathbf{v}_1, \mathbf{v}_4, \mathbf{v}_5\}, \{\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$$
$$\{\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_5\}, \{\mathbf{v}_2, \mathbf{v}_4, \mathbf{v}_5\}, \{\mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5\}.$$

It is possible to re-order the vectors in S so that, for any single choice of three linearly independent vectors, the columns, corresponding to your choice, are the pivot columns. Thus any one of these sets is a basis for V.

Now suppose that we wanted to choose 2 vectors from S and we want them to be linearly independent. Notice that it is not obvious by inspection whether or not two of these vectors are multiples of each other since we are using complex numbers. We would have to examine the system:

$$d_1\mathbf{v}_{q_1} + d_2\mathbf{v}_{q_2} = \mathbf{0}, \text{ where } 1 \leq q_1 \leq q_2 \leq 5.$$

And the coefficient matrix for this homogeneous linear system would consist of two columns of the matrix A. We need to arrange for each of these columns to be a pivot column. We could not, for example, choose $\mathbf{v}_1$ and $\mathbf{v}_2$.

Careful examination of $RREF(A)$ reveals that we can choose any pair of vectors except both $\mathbf{v}_1$ and $\mathbf{v}_2$. This gives a total of 9 different possible sets.

Finally, any subset of S consisting of one vector will be linearly independent, and the empty set is also linearly independent.

That gives a total of 22 different linearly independent subsets of S.

Theoretical questions involving linear independence and linear dependence.

The single most important piece of advice that we can provide when you are considering whether or not a set, $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ say, is either linearly dependent of linearly independent, is to begin by just writing down

$$\text{Consider } c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0},$$

and then use whatever additional information that you may have to process this equation.

Do not tell that this equation is true when $c_1 = c_2 = \cdots = c_p = 0$, because we all know this already.

Do not tell that this equation is only true when $c_1 = c_2 = \cdots = c_p = 0$, unless you know that S is linearly independent.

Lemma 6

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subseteq \mathbb{F}^n$, and suppose that S is linearly independent.
Let $\mathbf{w} \in \mathbb{F}^n$. The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{w}\}$ is linearly dependent **iff** $\mathbf{w} \in Span(S)$.

Proof

If $\mathbf{w} \in Span(S)$, then there exist scalars $c_1, c_2, \ldots, c_p \in \mathbb{F}$ such that

$$\mathbf{w} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p,$$

and then

$$\mathbf{w} - c_1\,\mathbf{v}_1 - c_2\,\mathbf{v}_2 - \cdots - c_p\,\mathbf{v}_p = \mathbf{0}.$$

Notice that the coefficient of $\mathbf{w}$ here is 1. We conclude that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{w}\}$ is linearly dependent.

On the other hand, suppose that the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p, \mathbf{w}\}$ is linearly dependent. Then there exist scalars $c_1, c_2, \ldots, c_p, d \in \mathbb{F}$, not all zero, such that

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p + d\,\mathbf{w} = \mathbf{0}.$$

There are two possibilities, either $d = 0$, and then

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p = \mathbf{0},$$

where not all these scalars are zero. This would mean that S is linearly dependent, which is a contradiction.

The other possibility is that $d \neq 0$, in which case, we can solve for $\mathbf{w}$ and write:

$$\mathbf{w} = \frac{-c_1}{d}\,\mathbf{v}_1 - \frac{c_2}{d}\,\mathbf{v}_2 - \cdots - \frac{c_p}{d}\,\mathbf{v}_p,$$

and in this case, we see that $\mathbf{w} \in Span(S)$. ■

Lemma 7

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subseteq \mathbb{F}^n$, and suppose that S is linearly independent.

Let $\mathbf{v}_k \in S$, Then $S \setminus \{\mathbf{v}_k\}$, is linearly independent .

Proof: left as an exercise.

Example 10

Let S_1 and S_2 be two linearly independent subsets of $\mathbb{F}^n$.

Let $V_1 = S_1 \cup S_2$ and $V_2 = S_1 \cap S_2$.

Are V_1 and V_2 linearly independent?

Solution

For the union, we might expect that this it too much to ask for, and
we can quickly construct a simple counterexample.

Let $S_1 = \left\{(1,0)^T\right\}, S_2 = \left\{(2,0)^T\right\}$ and $V_1 = S_1 \cup S_2 = \left\{(1,0)^T, (2,0)^T\right\}$,

then V_1 is clearly linearly dependent.

As for their intersection, this seems a little more reasonable. We can prove that it is true. We know that each of the two sets S_i have at most n vectors in them, by Corollary 1, and so V_2 has at most n vectors in it.

Let us write $V_2 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ with $p \leq n$. Now consider

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p = \mathbf{0}.$$

We recall that each of the vectors in this expression is in V_2, and thus each of these vectors is in S_1 (and S_2).
Since we know that S_1 is linearly independent, then we know that the only solution to the equation

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_p \mathbf{v}_p = \mathbf{0}$$

is the trivial linear combination, i.e. $c_1 = c_2 = \cdots = c_p = 0$, and so we can conclude that the set V_2 is linearly independent.

Topic 17C

Spanning and Bases

We have now got a good feel of how to determine whether or not a set is linearly independent, and if the set is linearly dependent then we know how to reduce it to a set that is linearly independent.

The second major property that a basis needs to possess is for it to be a spanning set. We begin this lecture with a discussion about spanning sets and then move on to the detection of a basis.

Suppose that we have a subspace V of $\mathbb{F}^n$, and suppose that S is a subset of V which has a finite number of elements in it. Thus $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, for some vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in V$.

Lemma 8

Let V be a subspace of $\mathbb{F}^n$ and let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subset V$, then $Span(S)$ is a subspace of V.

Proof

Let $\mathbf{v} \in Span(S)$. Then $\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p$, for some $c_1, c_2, \ldots, c_p \in \mathbb{F}$.

However, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in V$, and V is a subspace of $\mathbb{F}^n$, so we can use the inductive extension of the closure property of a subspace to conclude that $\mathbf{v} \in V$.

Note that we actually already know that $Span(S)$ is a subspace of $\mathbb{F}^n$ too, see Topic 17A example 1 (c).

Note that if we define a subspace of a subspace, in the natural way, then $Span(S)$ is a subspace of V. $\blacksquare$

Often we need to show that a subset S of V has the property that $Span(S) = V$. That is, we are trying to show that two sets are equal. The standard way to do this is to prove that each of them is contained in the other.

Lemma 8 tells us that if V is a subspace of $\mathbb{F}^n$, and if S is a subset of V, then we can immediately conclude that $Span(S) \subseteq V$, and thus the only property that needs investigation is whether or not $V \subseteq Span(S)$.

In order to prove that $V \subseteq Span(S)$, we need to show that if $\mathbf{v} \in V$, then $\mathbf{v} \in Span(S)$, that is, we need to show that there exist scalars $c_1, c_2, \ldots, c_p \in \mathbb{F}$ such that

$$\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p.$$

If we let $A = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$, the coefficient matrix of this system of equations, and if we let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p|\,\mathbf{v})$, the augmented matrix of this system of equations, then we need to show that $rank(A) = rank(B)$ for every $\mathbf{v} \in V$. Clearly, this will depend on both S and V.

Example 10

Consider the plane in $\mathbb{R}^3$ given by the scalar equation $2x - 6y + 4z = 0$.
We can solve this system of equations (i.e. this single equation), and let $z = t, y = s$, and then $x = 3s - 2t$.

We can then express the plane as $P = \left\{ \begin{pmatrix} 3s-2t \\ s \\ t \end{pmatrix} : s, t \in \mathbb{R} \right\}$.

Let $S_1 = \left\{ \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \\ 2 \end{pmatrix} \right\}$, does $Span(S_1) = P$?

Solution

Let $\mathbf{v} \in P$. Then we can write $\mathbf{v} = \begin{pmatrix} 3s_1-2t_1 \\ s_1 \\ t_1 \end{pmatrix}$ for some $s_1, t_1 \in \mathbb{R}$. Consider now,

$$\mathbf{v} = \begin{pmatrix} 3s_1-2t_1 \\ s_1 \\ t_1 \end{pmatrix} = a \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} + b \begin{pmatrix} -4 \\ 0 \\ 2 \end{pmatrix}, \quad \text{where } a, b \in \mathbb{R}.$$

The augmented matrix for this system is :

$$\begin{pmatrix} 2 & -4 & |\; 3s_1 - 2t_1 \\ 0 & 0 & |\; s_1 \\ -1 & 2 & |\; t_1 \end{pmatrix},$$

which row reduces to:

$$\begin{pmatrix} 2 & -4 & |\; 3s_1 - 2t_1 \\ 0 & 0 & |\; s_1 \\ 0 & 0 & |\; 0 \end{pmatrix}.$$

The last column of the augmented matrix is a pivot column (when $s_1 \neq 0$), and thus we cannot always solve this system. It is also quite clear from the form of the two vectors in S_1, which both have a middle component of zero, that you will be unable to build all the vectors in P from them, as some of the vectors in P have a non-zero middle component. Thus $Span(S_1) \neq P$.

We now consider $S_2 = \left\{ \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -4 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} \right\}$, does $Span(S_2) = P$?

Solution

Let $\mathbf{v} \in P$, $\mathbf{v} = \begin{pmatrix} 3s_1 - 2t_1 \\ s_1 \\ t_1 \end{pmatrix}$, for some $s_1, t_1 \in \mathbb{R}$. Consider now,

$$\mathbf{v} = \begin{pmatrix} 3s_1 - 2t_1 \\ s_1 \\ t_1 \end{pmatrix} = a \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} + b \begin{pmatrix} -4 \\ 0 \\ 2 \end{pmatrix} + c \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \quad \text{where } a, b, c \in \mathbb{R}.$$

The augmented matrix for this system is :

$$\left(\begin{array}{rrrl} 2 & -4 & 3 & \mid 3s_1 - 2t_1 \\ 0 & 0 & 1 & \mid s_1 \\ -1 & 2 & 0 & \mid t_1 \end{array} \right),$$

which row reduces to:

$$\left(\begin{array}{rrrl} 1 & -2 & 0 & \mid -t_1 \\ 0 & 0 & 1 & \mid s_1 \\ 0 & 0 & 0 & \mid 0 \end{array} \right).$$

We notice that the rank of the coefficient matrix and the rank of the augmented matrix are both equal to 2, and thus the system will be consistent for all $\mathbf{v} \in P$. That is, $Span(S_2) = P$.

In addition, we can solve the system to get $c = s_1, b = u$, and $a = 2u - t_1,$ for $u \in \mathbb{R}$. Note that the fact that there is a parameter u in this solution, tells us that we have redundancy. Thus, the set S_2 is not linearly independent.

We will have more examples of spanning sets for subspaces of $\mathbb{F}^n$ during the rest of the course. We continue, for the moment, by considering spanning sets for the whole of $\mathbb{F}^n$.

Lemma 9

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, for some vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{F}^n$. Then

$$Span(S) = \mathbb{F}^n \ \ \textbf{iff} \ \ rank((\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)) = n.$$

Proof

Consider the equation $\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p$.
Let $A = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)$ be the coefficient matrix and $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \,|\, \mathbf{v})$, be the augmented matrix with $\mathbf{v} \in \mathbb{F}^n$.

We know from our discussion above, that:

$$\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_p\,\mathbf{v}_p \text{ has a solution } \textbf{iff } rank(A) = rank(B).$$

We now want this equation to have a solution for any $\mathbf{v} \in \mathbb{F}^n$, and

this happens **iff** the last column of B cannot be a pivot column, and

this happens **iff** each **row** of A has a pivot in it, and A has n rows. ■

Note that if $rank((\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p)) = n$, then $p \geq n$, and so this result is very reasonable.
We need at least n vectors in S for the $Span(S)$ to be equal to $\mathbb{F}^n$.
However the vectors in S must be appropriate.

Example 11

(a) Let $S_1 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 3 \\ -3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix} \right\}$. Does S_1 span $\mathbb{R}^4$?

Solution

No, there are not enough vectors, we need at least 4 vectors to span $\mathbb{R}^4$.

(b) Let $S_2 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 3 \\ -3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}, \begin{pmatrix} 5 \\ 3 \\ 10 \\ 6 \end{pmatrix} \right\}$. Does S_2 span $\mathbb{R}^4$?

Solution

Perhaps: let us see if these four vectors span $\mathbb{R}^4$. Let $(\alpha, \beta, \gamma, \delta)^T \in \mathbb{R}^4$, let us consider

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ -2 \\ 3 \\ -3 \end{pmatrix} + c_3 \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix} + c_4 \begin{pmatrix} 5 \\ 3 \\ 10 \\ 6 \end{pmatrix},$$

for which the coefficient matrix is:

$$\begin{pmatrix} 1 & 2 & 2 & 5 \\ 1 & -2 & 4 & 3 \\ 1 & 3 & 6 & 10 \\ 1 & -3 & 8 & 6 \end{pmatrix}.$$

The reduced row echelon form of this is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since we do not have 4 pivots, we conclude the $Span(S_2) \neq \mathbb{R}^4$.

Why is there a problem? We have four vectors, but they are linearly dependent.
Were we to consider the equation

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ -2 \\ 3 \\ -3 \end{pmatrix} + c_3 \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix} + c_4 \begin{pmatrix} 5 \\ 3 \\ 10 \\ 6 \end{pmatrix},$$

we would have non-trivial solutions. In fact, we can see that the vector which we added, $(5, 3, 10, 6)^T$ lies in $Span(S_1)$, and thus this new vector is of no use to us at all.

(c) Let $S_3 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 3 \\ -3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}, \begin{pmatrix} 5 \\ 3 \\ 10 \\ 6 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \\ 6 \\ 10 \end{pmatrix} \right\}$. Does S_3 span $\mathbb{R}^4$?

Perhaps: let us see if these five vectors span $\mathbb{R}^4$. Let $(\alpha, \beta, \gamma, \delta)^T \in \mathbb{R}^4$. Let us consider

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ -2 \\ 3 \\ -3 \end{pmatrix} + c_3 \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix} + c_4 \begin{pmatrix} 5 \\ 3 \\ 10 \\ 6 \end{pmatrix} + c_5 \begin{pmatrix} 3 \\ 5 \\ 6 \\ 10 \end{pmatrix},$$

for which the coefficient matrix is:

$$\begin{pmatrix} 1 & 2 & 2 & 5 & 3 \\ 1 & -2 & 4 & 3 & 5 \\ 1 & 3 & 6 & 10 & 6 \\ 1 & -3 & 8 & 6 & 10 \end{pmatrix}.$$

The reduced row echelon form of this is

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Since we have a pivot in each of the 4 rows of the coefficient matrix, we can conclude that the system of equations has a solution for any $(\alpha, \beta, \gamma, \delta)^T \in \mathbb{R}^4$. Thus $Span(S_3) = \mathbb{R}^4$.

Basis

In Topic 17A, we defined a basis for a subspace to be a subset of that subspace which has the two additional properties that it is a linearly independent subset, and that this subset spans the subspace.

We will first consider the idea of a basis for the space $\mathbb{F}^n$, and in future lectures, we will consider bases for subspaces as they arise.

Lemma 10

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, for some vector $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{F}^n$. Then

if S is a basis for $\mathbb{F}^n$, then S has exactly n vectors, that is, $p = n$.

Proof

For the subset S of $\mathbb{F}^n$ to be a basis of $\mathbb{F}^n$, then we must have that:

(i) $Span(S) = \mathbb{F}^n$, and from Lemma 9, this means that $p \geq n$, and

(ii) S is linearly independent, and from Corollary 1 (T17B), this means that $p \leq n$.

We conclude that $p = n$. ■

If we have a subset, S, of $\mathbb{F}^n$, which has less than n vectors in it, then $Span(S) \neq \mathbb{F}^n$: we do not have enough vectors in the set S to build the entire space $\mathbb{F}^n$.

If we have a subset, S, of $\mathbb{F}^n$, which has more than n vectors in it, then S must be linearly dependent, we have redundancy. Note that we *may* not have the **appropriate** vectors in the set S to build the entire space $\mathbb{F}^n$.

If we have a subset, S, of $\mathbb{F}^n$, which has exactly n vectors in it, then there is a possibility that S is a basis for V, we will have to check the set S for linear independence and for spanning.

The following result tells us that we need only check for one of these properties.

Lemma 11

$S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, for n distinct vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{F}^n$. Then

$$S \text{ is linearly independent } \textbf{ iff } \ Span(S) = \mathbb{F}^n.$$

Proof

When we are checking the set S for linear independence, we must consider the equation:

$$c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n = \mathbf{0}.$$

If we let the coefficient matrix for this system be the $(n \times n)$ matrix $A = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. S is linearly independent **iff** $rank(A) = n$.

When we are checking to see whether or not $Span(S) = \mathbb{F}^n$, we must consider the equation:

$$\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n,$$

where $\mathbf{v}$ is an arbitrary vector in $\mathbb{F}^n$. The coefficient matrix for this system is A, from above. The equation has a solution for any $\mathbf{v} \in \mathbb{F}^n$ **iff** A has a pivot in each row **iff** $rank(A) = n$. Thus $Span(S) = \mathbb{F}^n$ **iff** $rank(A) = n$.

Putting all this together yields:
S is linearly independent **iff** $rank(A) = n$ **iff** $Span(S) = \mathbb{F}^n$. ■

Thus when we are checking to see whether or not a subset, S, of $\mathbb{F}^n$ is a basis for $\mathbb{F}^n$, there are three ways to proceed:

(a) Check S for linear independence, and check S for spanning, or

(b) Count the vectors in S and check S for spanning, or

(c) Count the vectors in S and check S for linear independence.

Option (c) is usually to be the quickest.

Example 12

Let $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_5 = \begin{pmatrix} 4 \\ -3 \\ 2 \\ -1 \end{pmatrix},$

Which of the following subsets of $\mathbb{R}^4$ is a basis for $\mathbb{R}^4$?

$$A = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}\,,\ B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5\}\,,\ C = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}\,,\ D = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_5\}\,.$$

Solution

Sets A and B fail immediately as they have the wrong number of vectors in them.

Sets C and D have the correct number of vectors in them so we need to investigate them further, that is, we need to check them for linear independence.

For set C, we consider the equation $c_1\,\mathbf{v_1} + c_2\,\mathbf{v}_2 + c_3\,\mathbf{v}_3 + c_4\,\mathbf{v}_4 = \mathbf{0}$. The coefficient matrix is:

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

which row reduces to:

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{5} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{5} \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since this matrix has only three pivots, then the four vectors are linearly dependent and so the set C is not a basis for $\mathbb{R}^4$.

For set D, we consider $c_1\,\mathbf{v_1} + c_2\,\mathbf{v}_2 + c_3\,\mathbf{v}_3 + c_5\,\mathbf{v}_5 = \mathbf{0}$. The coefficient matrix is:

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

which row reduces to:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Since we have four pivots, we conclude that D is a linearly independent subset.

Thus D is a basis for $\mathbb{R}^4$.

Definition 5: Dimension

We refer to the number n, which is equal to the number of elements in a basis for $\mathbb{F}^n$, as the **dimension** of $\mathbb{F}^n$, and we write $dim(\mathbb{F}^n) = n$.

We also say that $\mathbb{F}^n$ is n-**dimensional.**

The **dimension** of a space is equal to the number of elements in a basis for that space.

Definition 6: Standard basis

The **standard basis** for $\mathbb{F}^n$ is the set of n vectors, $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ with

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ with the 1 in the } i^{th} \text{ row}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

There are two natural problems that you might encounter when trying to obtain a basis for $\mathbb{F}^n$.

(a) You have a subset of vectors, S, in $\mathbb{F}^n$ with the property that $Span(S) = \mathbb{F}^n$. However, you know that you have too many vectors for a basis, and so S is linearly dependent. Lemma 5 (T17B) indicates to us how to reduce this set to a linearly independent subset with the same span. If we follow the guidelines of that lemma, then we will produce a subset of S which is a basis for $\mathbb{F}^n$.

(b) There is a subset, U, of $r < n$ linearly independent vectors, which you prefer, but you do not have enough vectors for a basis. In this case, we would need to add $(n - r)$ vectors to the set in order to obtain a basis for $\mathbb{F}^n$. You usually do not know which vectors to add. A natural choice is to add the entire n vectors of the standard basis to your set U, to produce the set S. Now $Span(S) = \mathbb{F}^n$, but you have r vectors too many , and so we are back to a). As long as you add the new (n) vectors **after** the (r) vectors that you started with, then Lemma 5 (T17B) will produce a basis for you which will contain the subset, U, of r vectors which you prefer.
We will encounter examples of both these types as we continue through the course.

We conclude with the most important result about a basis. This result tells us why a basis is important and opens the door to the idea of components and coordinates.

Theorem 1: Unique Representation Theorem.

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{F}^n$, then, for every vector $\mathbf{v} \in \mathbb{F}^n$, there exist unique scalars $c_1, c_2, \ldots, c_n \in \mathbb{F}$ such that $\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n$.

Proof

Existence: since B is a basis for $\mathbb{F}^n$, then $Span(B) = \mathbb{F}^n$, and so the vector $\mathbf{v}$ can be written as a linear combination of elements of B, that is, there exist scalars $c_1, c_2, \ldots, c_n \in \mathbb{F}$ such that $\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n$.

Uniqueness: suppose that $\mathbf{v} = d_1\,\mathbf{v}_1 + d_2\,\mathbf{v}_2 + \cdots + d_n\,\mathbf{v}_n$, for some scalars $d_1, d_2, \ldots, d_n \in \mathbb{F}$. If we subtract these two expressions for $\mathbf{v}$, then we have:

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n - (d_1\,\mathbf{v}_1 + d_2\,\mathbf{v}_2 + \cdots + d_n\,\mathbf{v}_n)$$

$$\mathbf{0} = (c_1 - d_1)\mathbf{v}_1 + (c_2 - d_2)\mathbf{v}_2 + \cdots + (c_n - d_n)\mathbf{v}_n.$$

As B is linearly independent, we conclude that $(c_1 - d_1) = 0, (c_2 - d_2){=}0, \ldots, (c_n - d_n) = 0$.

That is, $(c_1 = d_1), (c_2 = d_2), \ldots, (c_n = d_n)$, and thus the representation for $\mathbf{v}$ is unique. ■

Topic 17D

Bases, Coordinates and Components

Theorem 1: Unique Representation Theorem (From Topic 17C).

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{F}^n$, then, for every vector $\mathbf{v} \in \mathbb{F}^n$, there exist unique scalars $c_1, c_2, \ldots, c_n \in \mathbb{F}$ such that $\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n$.

Definition 7: Coordinates and components

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{F}^n$. Let the vector $\mathbf{v} \in \mathbb{F}^n$ with

$$\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n = \sum_{i=1}^{n} c_i\,\mathbf{v}_i.$$

Then we use the term **coordinates** and **components** to mean the scalars $c_1, c_2, \ldots, c_n$.

We use the term **coordinate vector** (often just coordinates) and **component vector** (often just components) to refer to the column vector in $\mathbb{F}^n$, constructed from these scalars c_i, **where the ordering of these scalars c_i's matches the order in which the vectors $\mathbf{v}_i$ appear in the basis B.**
We denote this object by $[\mathbf{v}]_B$. That is,

$$[\mathbf{v}]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

Note that this relationship is a two-way relationship.

If $\mathbf{v} \in \mathbb{F}^n$, then $\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n = \sum_{i=1}^{n} c_i\mathbf{v}_i$, and so, $[\mathbf{v}]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$,

that is, if you have a vector in $\mathbb{F}^n$, and if you have a basis, B, then you can find the components (see examples later in this lecture).

Also, if $[\mathbf{v}]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$, then $\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n$, that is, if you have the component vector of $\mathbf{v}$ and if you know the basis, B, then you can write down the vector $\mathbf{v}$.

We think of the basis, B, as the key for coding: from $\mathbf{v}$ to $[\mathbf{v}]_B$, and from $[\mathbf{v}]_B$ to $\mathbf{v}$.

Note: in linear algebra we **often** use the standard basis, denoted by S. In fact, in the early parts of the course, and really, right up to this point, this is the only basis we have any notion of, without really knowing what a basis is! We used this standard basis without even knowing that we are using it, and we use it to provide one of our two standard ways of actually communicating with each other what our vector is. The second way was the loose geometric description, which we always suspect is just not as precise and thus not as useful.

We have been writing expressions such as

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \ (*)$$

when we should really write

$$[\mathbf{v}]_S = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}. \ (**)$$

We are being rather lazy and careless, when we write the former $(*)$ instead of the later $(**)$, however we all understand what we mean by this and we will continue to do it. Whenever someone tells us that

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}, \ (*)$$

then we understand that we really mean that

$$[\mathbf{v}]_S = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \Leftrightarrow \mathbf{v} = v_1\,\mathbf{e}_1 + v_2\,\mathbf{e}_2 + \cdots + v_n\,\mathbf{e}_n = \sum_{i=1}^{n} v_i\,\mathbf{e}_i.$$

Moving forwards, when we communicate with each other about some vector in $\mathbb{F}^n$, we still have to represent it somehow, we must still write it down, and, unless we have some other arrangement then we will still be doing this by making us of the standard basis.

Notation

The notation of $[\mathbf{v}]_S$, should remind you of $[T]_S$, for the standard matrix representation of the linear transformation, T, and there is a very good reason why they are very similar: the three parts to $[\mathbf{v}]_B$ are: the square brackets [ ] which reminds us that we are taking coordinates, that is, we are representing something with an array (column vector) of numbers, B, which reminds us that we are using the basis B, and, of course, $\mathbf{v}$, to indicate the object (vector) whose coordinates are being taken.

We will now have many examples of using bases other than the standard basis, most often, the vectors that you will be given, are written down with respect to the standard basis.

Example 13

Show that $B = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 6 \end{pmatrix} \right\}$ is a basis for $\mathbb{R}^2$. Find the coordinates of $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ in the basis B, and if $[\mathbf{w}]_B = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$, then find $[\mathbf{w}]_S$.

Solution

Note, as usual, all of the supplied vectors have been written with respect to the standard basis. Otherwise we could not write them down.

We have two vectors in two dimensions. We need to check if they are linearly independent and spanning $\mathbb{R}^2$, and this can be done by evaluating the rank of

$$A = \begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix}.$$

Row reduction yields the matrix $R = \begin{pmatrix} 1 & 5 \\ 0 & -4 \end{pmatrix}$, which has rank 2.
Thus we have a basis.

It is then possible to write $\begin{pmatrix} 3 \\ 4 \end{pmatrix} = a \begin{pmatrix} 1 \\ 2 \end{pmatrix} + b \begin{pmatrix} 5 \\ 6 \end{pmatrix}$, for some a and b in $\mathbb{R}$.

We need to find the components a and b. Consider the system

$$\begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix},$$

which has augmented matrix

$$\left(\begin{array}{cc|c} 1 & 5 & 3 \\ 2 & 6 & 4 \end{array} \right) \text{ which reduces to } \left(\begin{array}{cc|c} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{array} \right).$$

Thus $a = \frac{1}{2}$ and $b = \frac{1}{2}$, and we have

$$\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 5 \\ 6 \end{pmatrix} \quad \text{and} \quad \left[\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right]_B = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Secondly, we are given that

$$[\mathbf{w}]_B = \begin{pmatrix} -3 \\ 2 \end{pmatrix},$$

this means that :

$$\mathbf{w} = (-3)\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}, \text{ that is,}$$

$$[\mathbf{w}]_S = \mathbf{w} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}.$$

Example 14

Show that $B = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix} \right\}$ is a basis for $\mathbb{R}^3$. Find the coordinates of $\begin{pmatrix} -3 \\ 2 \\ -6 \end{pmatrix}$ in this basis, and if $[\mathbf{w}]_B = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}$, then obtain $[\mathbf{w}]_S$.

Solution

We have three vectors in three dimensions. We need to check if they are linearly independent and spanning $\mathbb{R}^3$, and this can be done by evaluating the rank of

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

Row reduction yields the matrix $R = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix}$, which has a rank 3.

Thus we have a basis.

It is then possible to write $\begin{pmatrix} -3 \\ 2 \\ -6 \end{pmatrix} = a\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + b\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} + c\begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix}$, for some $a, b, c \in \mathbb{R}$.

We need to find the components a, b and c. Consider the system:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -3 \\ 2 \\ -6 \end{pmatrix},$$

which has augmented matrix

$$\begin{pmatrix} 1 & 1 & 1 & | -3 \\ 1 & 0 & -2 & | \, 2 \\ 1 & 3 & 4 & | -6 \end{pmatrix} \text{ which reduces to } \begin{pmatrix} 1 & 0 & 0 & | \frac{-8}{3} \\ 0 & 1 & 0 & | \, 2 \\ 0 & 0 & 1 & | \frac{-7}{3} \end{pmatrix}.$$

Thus $a = -\frac{8}{3}$, $b = 2$ and $c = -\frac{7}{3}$, and we have

$$\begin{pmatrix} -3 \\ 2 \\ -6 \end{pmatrix} = \frac{-8}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} - \frac{7}{3} \begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix}, \quad \text{and}$$

$$\begin{bmatrix} -3 \\ 2 \\ -6 \end{bmatrix}_B = \begin{pmatrix} -\frac{8}{3} \\ 2 \\ -\frac{7}{3} \end{pmatrix}.$$

Secondly, we are given that

$$[\mathbf{w}]_B = \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix},$$

this means that

$$\mathbf{w} = (3) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (-4) \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} + (5) \begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ -7 \\ 11 \end{pmatrix}, \text{ that is,}$$

$$[\mathbf{w}]_S = \mathbf{w} = \begin{pmatrix} 4 \\ -7 \\ 11 \end{pmatrix}.$$

Suppose we now tell you that there are some more vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, which I have been given in the standard basis, but whose coordinates I would like to obtain in the non-standard basis. Do we really have to do the entire calculation again and again? The answer is no.

Let us revisit Example 13 to see what the actual performed calculations involved.

Example 15

Let us first of all consider the end of the calculation in Example 13, when we had

$$[\mathbf{w}]_B = \begin{pmatrix} -3 \\ 2 \end{pmatrix}, \text{ and then } [\mathbf{w}]_S = \mathbf{w} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}.$$

This was completed by performing the calculation

$$\mathbf{w} = (-3)\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} -3 \\ 2 \end{pmatrix},$$

that is, there is a matrix at work here, $M_1 = \begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix}$, and it will change coordinates from the new system (basis B) to the standard system (basis S), and it will do this for any vector whose components we already have in B.

Notice that M_1 is very simple, and special: its columns are the coordinates of the new basis vector in the standard basis.

If you now tell me that you have a vector $\mathbf{z}$, with

$$[\mathbf{z}]_B = \begin{pmatrix} 4 \\ -3 \end{pmatrix}, \text{ then } [\mathbf{z}]_S = \mathbf{z} = \begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} 4 \\ -3 \end{pmatrix} = \begin{pmatrix} -11 \\ -10 \end{pmatrix}.$$

What about going the other way? Well, lets us look at what we did in the first part of the Example 13.

We wanted to obtain $\left[\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right]_B$, and we need to solve the system:

$$\begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

Notice that the coefficient matrix here is the invertible matrix M_1, and we can write:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 5 \\ 2 & 6 \end{pmatrix}^{-1}\begin{pmatrix} 3 \\ 4 \end{pmatrix} = -\frac{1}{4}\begin{pmatrix} 6 & -5 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix},$$

and thus yielding the result that $\left[\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right]_B = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$.

Once again there is a matrix at work here, $M_1^{-1} = -\frac{1}{4}\begin{pmatrix} 6 & -5 \\ -2 & 1 \end{pmatrix}$, and it will change coordinates from the standard system (basis S) to the new system (basis B), and it will do this for any vector whose components we already have in S.

Notice that M_1^{-1} is very simple, and easily obtained, it is just the inverse matrix of the matrix whose columns are the coordinates of the new basis vector in the standard basis. With a little thought, you can see that the columns of M_1^{-1} are the coordinates of the standard basis vectors in the new basis.

If you now tell me that you have a vector $\mathbf{p}$, with

$$\mathbf{p} = [\mathbf{p}]_S = \begin{pmatrix} -2 \\ 5 \end{pmatrix}, \text{ then } [\mathbf{p}]_B = -\frac{1}{4}\begin{pmatrix} 6 & -5 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} -2 \\ 5 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 37 \\ -9 \end{pmatrix}.$$

Lemma 12: Taking coordinates is a linear transformation.

Let B be a basis for $\mathbb{F}^n$. Then $[\ ]_B : \mathbb{F}^n \to \mathbb{F}^n$ given by $\mathbf{x} \mapsto [\mathbf{x}]_B$ is a linear transformation.

Proof

We implicitly built this into the operation of taking coordinates in Topic 1. ■

We make use of this result to prove the following lemma which summarizes much of what we have learned in this lecture.

Lemma 13

Let $B_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $B_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ be bases for $\mathbb{F}^n$.

Let $\mathbf{x} \in \mathbb{F}^n$ with $[\mathbf{x}]_{B_1} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ and $[\mathbf{x}]_{B_2} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$.

Then $[\mathbf{x}]_{B_2} = {}_{B_2}[I]_{B_1}\ [\mathbf{x}]_{B_1}$ and $[\mathbf{x}]_{B_1} = {}_{B_1}[I]_{B_2}\ [\mathbf{x}]_{B_2}$, where

$${}_{B_2}[I]_{B_1} = \left([\mathbf{v}_1]_{B_2}, [\mathbf{v}_2]_{B_2}, \ldots, [\mathbf{v}_n]_{B_2}\right)$$
and
$${}_{B_1}[I]_{B_2} = \left([\mathbf{w}_1]_{B_1}, [\mathbf{w}_2]_{B_1}, \ldots, [\mathbf{w}_n]_{B_1}\right).$$

Proof

Suppose $[\mathbf{x}]_{B_1} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$, then $\mathbf{x} = a_1\,\mathbf{v}_1 + a_2\,\mathbf{v}_2 + \cdots + a_n\,\mathbf{v}_n$ and so

$$
\begin{aligned}
[\mathbf{x}]_{B_2} &= [a_1\,\mathbf{v}_1 + a_2\,\mathbf{v}_2 + \cdots + a_n\,\mathbf{v}_n]_{B_2} \\
&= a_1\,[\mathbf{v}_1]_{B_2} + a_2\,[\mathbf{v}_2]_{B_2} + \cdots + a_n\,[\mathbf{v}_n]_{B_2}\,, \text{ as taking coordinates is a linear operation,} \\
&= \left([\mathbf{v}_1]_{B_2}\,, [\mathbf{v}_2]_{B_2}\,, \ldots, [\mathbf{v}_n]_{B_2}\right) \begin{pmatrix} a_1 \\ a_n \\ \vdots \\ a_n \end{pmatrix} = {}_{B_2}[I]_{B_1}\;\; [\mathbf{x}]_{B_1}\,.
\end{aligned}
$$

The proof going from B_2 to B_1 is identical, with B_2 switched with B_1. ■

Definition 8: Change-of-coordinate matrix, change-of-basis matrix.

We refer to the matrix, ${}_{B_2}[I]_{B_1}$, as the **change-of-coordinate matrix** from B_1-coordinates to B_2-coordinates.

It is also called the **change-of-basis matrix** from basis B_1 to basis B_2.

We refer to the matrix, ${}_{B_1}[I]_{B_2}$ as the change-of-coordinate matrix from B_2-coordinates to B_1-coordinates.

It is also called the **change-of-basis matrix** from basis B_2 to basis B_1.

The role of the change-of-coordinate matrix coordinates is exactly as its name suggests: it changes coordinates from one basis to another.

Corollary 2

Let $B_1 = S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, the standard basis for $\mathbb{F}^n$, and let $B_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ be another basis for $\mathbb{F}^n$.

Let $\mathbf{x} \in \mathbb{F}^n$ with $[\mathbf{x}]_S = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ and $[\mathbf{x}]_{B_2} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$.

Then $[\mathbf{x}]_{B_2} = {}_{B_2}[I]_S\;\; [\mathbf{x}]_S$ and $[\mathbf{x}]_S = {}_S[I]_{B_2}\;\; [\mathbf{x}]_{B_2}$, where

$$
{}_{B_2}[I]_S = \left([\mathbf{e}_1]_{B_2}\,, [\mathbf{e}_2]_{B_2}\,, \ldots, [\mathbf{e}_n]_{B_2}\right)
$$

and

$$
{}_S[I]_{B_2} = ([\mathbf{w}_1]_S\,, [\mathbf{w}_2]_S\,, \ldots, [\mathbf{w}_n]_S) = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)\,.
$$

The proof is immediate from Lemma 13, on letting $\mathbf{v}_i = \mathbf{e}_i$, for $i = 1, \ldots, n$. ■

Notice that in this case ${}_S[I]_{B_2}$ is easily obtained. However you usually want to go the other way, and thus require, ${}_{B_2}[I]_S$.

The next corollary tells us that once we have one of these two matrices then we can obtain the other rather quickly as they are inverses of each other.

Corollary 3

The two change-of-basis matrices ${}_{B_1}[I]_{B_2}$ and ${}_{B_2}[I]_{B_1}$ are inverses of each other, that is,

$$ {}_{B_1}[I]_{B_2}\ {}_{B_2}[I]_{B_1} = I_n \quad \text{and} \quad {}_{B_2}[I]_{B_1}\ {}_{B_1}[I]_{B_2} = I_n. $$

Proof

Suppose we change coordinates twice: we start in basis B_1, then change basis to B_2, and then back to B_1.

We then have, for any $\mathbf{x} \in \mathbb{F}^n$,

$$ \begin{aligned} [\mathbf{x}]_{B_2} &= {}_{B_2}[I]_{B_1} \quad [\mathbf{x}]_{B_1} \quad \text{and} \\ [\mathbf{x}]_{B_1} &= {}_{B_1}[I]_{B_2} \quad [\mathbf{x}]_{B_2} = {}_{B_1}[I]_{B_2} \quad {}_{B_2}[I]_{B_1} \quad [\mathbf{x}]_{B_1}. \end{aligned} $$

We can write this as

$$ (I_n - {}_{B_1}[I]_{B_2} \quad {}_{B_2}[I]_{B_1})\ [\mathbf{x}]_{B_1} = \mathbf{0}, \text{ for any } \ \mathbf{x} \in \mathbb{F}^n. $$

We can conclude that

$$ (I_n - {}_{B_1}[I]_{B_2} \quad {}_{B_2}[I]_{B_1}) = \mathbb{O}, \quad \text{the zero matrix} $$

and thus ${}_{B_1}[I]_{B_2}$ and ${}_{B_2}[I]_{B_1}$ are inverses of each other. ■

Example 16

Let us use this new information to revisit our earlier Example 14, where we had

$$ {}_S[I]_B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & 3 & 4 \end{pmatrix}. $$

We invert this matrix to find that

$$ {}_B[I]_S = \tfrac{1}{3} \begin{pmatrix} 6 & -1 & -2 \\ -6 & 3 & 3 \\ 3 & -2 & -1 \end{pmatrix} \text{ and so} $$

$$\left[\begin{pmatrix} -3 \\ 2 \\ -6 \end{pmatrix}\right]_B = \tfrac{1}{3}\begin{pmatrix} 6 & -1 & -2 \\ -6 & 3 & 3 \\ 3 & -2 & -1 \end{pmatrix}\begin{pmatrix} -3 \\ 2 \\ -6 \end{pmatrix} = \tfrac{1}{3}\begin{pmatrix} -8 \\ 6 \\ -7 \end{pmatrix}.$$

Furthermore, if we now want, for example, $\left[\begin{pmatrix} 3 \\ 5 \\ 9 \end{pmatrix}\right]_B$, then we evaluate:

$$\left[\begin{pmatrix} 3 \\ 5 \\ 9 \end{pmatrix}\right]_B = \frac{1}{3}\begin{pmatrix} 6 & -1 & -2 \\ -6 & 3 & 3 \\ 3 & -2 & -1 \end{pmatrix}\begin{pmatrix} 3 \\ 5 \\ 9 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -5 \\ 24 \\ -10 \end{pmatrix}.$$

Example 17

Suppose we are in $\mathbb{C}^2$ and we wish to introduce a new basis $B_2 = \{\mathbf{v}_1, \mathbf{v}_2\}$ with $\mathbf{v}_1 = \begin{pmatrix} 1+2i \\ 3+4i \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -1-4i \\ 3-2i \end{pmatrix}$.

Let $\mathbf{z} = \begin{pmatrix} 5-2i \\ 6-4i \end{pmatrix}$. Find the coordinates of $\mathbf{z}$ in B_2, $[\mathbf{z}]_{B_2}$.

Solution

Note that, as usual, all the vectors have been given with respect to the standard basis, this is our natural way to write them down.

$$_S[I]_{B_2} = \begin{pmatrix} 1+2i & -1-4i \\ 3+4i & 3-2i \end{pmatrix}, \text{ and}$$

$$_{B_2}[I]_S = \begin{pmatrix} 1+2i & -1-4i \\ 3+4i & 3-2i \end{pmatrix}^{-1} = \frac{1}{-6+20i}\begin{pmatrix} 3-2i & 1+4i \\ -3-4i & 1+2i \end{pmatrix}$$

$$= \frac{-6-20i}{436}\begin{pmatrix} 3-2i & 1+4i \\ -3-4i & 1+2i \end{pmatrix} = \frac{-1}{218}\begin{pmatrix} 29+24i & -37+22i \\ 31-42i & -17+16i \end{pmatrix}.$$

And we can now obtain $[\mathbf{z}]_{B_2}$ by

$$[\mathbf{z}]_{B_2} = {}_{B_2}[I]_S \; [\mathbf{z}]_S = \frac{-1}{218}\begin{pmatrix} 29+24i & -37+22i \\ 31-42i & -17+16i \end{pmatrix}\begin{pmatrix} 5-2i \\ 6-4i \end{pmatrix} = \frac{-1}{218}\begin{pmatrix} 59+342i \\ 33-108i \end{pmatrix}.$$

On the other hand, suppose we had a vector $\mathbf{w}$, with $[\mathbf{w}]_{B_2} = \begin{pmatrix} 2 \\ 3i \end{pmatrix}$.

We then have

$$[\mathbf{w}]_S = \mathbf{w} = {}_S[I]_{B_2} \; [\mathbf{w}]_{B_2} = \begin{pmatrix} 1+2i & -1-4i \\ 3+4i & 3-2i \end{pmatrix}\begin{pmatrix} 2 \\ 3i \end{pmatrix} = \begin{pmatrix} 14+i \\ 12+17i \end{pmatrix}.$$

Topic 17E

Examples

Let us start this topic with more examples of making use of unusual bases.

Example 18

Find the coordinates of $\mathbf{w} = \begin{pmatrix} 3 \\ 6 \\ 4 \end{pmatrix}$ in the basis $B = \left\{ \begin{pmatrix} 1 \\ 8 \\ 10 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \right\}$,

in two ways: one way, using the definition of coordinates, and the other way, using the change of basis matrix.

Solution

Method 1 - using the definition of coordinates: that is, find the constants, a, b, c such that:

$$\mathbf{w} = \begin{pmatrix} 3 \\ 6 \\ 4 \end{pmatrix} = a \begin{pmatrix} 1 \\ 8 \\ 10 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}.$$

The augmented matrix for this system is:

$$\left(\begin{array}{ccc|c} 1 & 1 & 4 & 3 \\ 8 & 1 & 5 & 6 \\ 10 & 1 & 6 & 4 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 4 & 3 \\ 0 & -7 & -27 & -18 \\ 0 & -9 & -34 & -26 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 4 & 3 \\ 0 & 1 & \frac{27}{7} & \frac{18}{7} \\ 0 & -9 & -34 & -7 \end{array} \right) \longrightarrow$$

$$\left(\begin{array}{ccc|c} 1 & 1 & 4 & 3 \\ 0 & 1 & \frac{27}{7} & \frac{18}{7} \\ 0 & 0 & \frac{5}{7} & \frac{-20}{7} \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 4 & 3 \\ 0 & 1 & \frac{27}{7} & \frac{6}{7} \\ 0 & 0 & 1 & -4 \end{array} \right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & 19 \\ 0 & 1 & 0 & 18 \\ 0 & 0 & 1 & -4 \end{array} \right) \longrightarrow$$

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 18 \\ 0 & 0 & 1 & -4 \end{array} \right).$$

We conclude that $a = 1$, $b = 18$, and $c = -4$: and thus, $[\mathbf{w}]_B = \begin{pmatrix} 1 \\ 18 \\ -4 \end{pmatrix}$.

Method 2 - we find the change of basis matrix:

$$ {}_S[I]_B = \begin{pmatrix} 1 & 1 & 4 \\ 8 & 1 & 5 \\ 10 & 1 & 6 \end{pmatrix}, $$

and we need to invert this matrix.

$$ \left(\begin{array}{ccc|ccc} 1 & 1 & 4 & 1 & 0 & 0 \\ 8 & 1 & 5 & 0 & 1 & 0 \\ 10 & 1 & 6 & 0 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 4 & 1 & 0 & 0 \\ 0 & -7 & -27 & -8 & 1 & 0 \\ 0 & -9 & -34 & -10 & 0 & 1 \end{array}\right) \longrightarrow $$

$$ \left(\begin{array}{ccc|ccc} 1 & 3 & 4 & 1 & 0 & 0 \\ 0 & 1 & \frac{27}{7} & \frac{8}{7} & \frac{-1}{7} & 0 \\ 0 & 0 & \frac{5}{7} & \frac{2}{7} & \frac{-9}{7} & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 3 & 4 & 1 & 0 & 0 \\ 0 & 1 & \frac{27}{22} & \frac{8}{22} & \frac{-1}{22} & 0 \\ 0 & 0 & 1 & \frac{2}{5} & \frac{-9}{5} & \frac{7}{5} \end{array}\right) \longrightarrow $$

$$ \left(\begin{array}{ccc|ccc} 1 & 3 & 0 & \frac{-3}{5} & \frac{36}{5} & \frac{-28}{5} \\ 0 & 1 & 0 & \frac{-2}{5} & \frac{34}{5} & \frac{-27}{5} \\ 0 & 0 & 1 & \frac{2}{5} & \frac{-9}{5} & \frac{7}{5} \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{-1}{5} & \frac{2}{5} & \frac{-1}{5} \\ 0 & 1 & 0 & \frac{-2}{5} & \frac{34}{5} & \frac{-27}{5} \\ 0 & 0 & 1 & \frac{2}{5} & \frac{-9}{5} & \frac{7}{5} \end{array}\right). $$

We conclude that

$$ {}_B[I]_S = \frac{1}{5} \begin{pmatrix} -1 & 2 & -1 \\ -2 & 34 & -27 \\ 2 & -9 & 7 \end{pmatrix}, $$

and so we have

$$ [\mathbf{w}]_B = \frac{1}{5} \begin{pmatrix} -1 & 2 & -1 \\ -2 & 34 & -27 \\ 2 & -9 & 7 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \\ 4 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 5 \\ 90 \\ -20 \end{pmatrix} = \begin{pmatrix} 1 \\ 18 \\ -4 \end{pmatrix}. $$

Example 19

Consider the three bases in $\mathbb{R}^2$:

$$ S\,,\ B_1 = \{\mathbf{v}_1, \mathbf{v}_2\} = \left\{ \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \begin{pmatrix} -3 \\ 2 \end{pmatrix} \right\},\ B_2 = \{\mathbf{w}_1, \mathbf{w}_2\} = \left\{ \begin{pmatrix} 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ 8 \end{pmatrix} \right\}. $$

(a) Find ${}_{B_1}[I]_{B_2}$ in two different ways, one direct and one making use of the standard basis.

(b) Write down ${}_{B_2}[I]_{B_1}$.

(c) Make use of these matrices to obtain $[\mathbf{z}]_{B_1}$ when $[\mathbf{z}]_{B_2} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$,

and to obtain $[\mathbf{p}]_{B_2}$ when $[\mathbf{p}]_{B_1} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$.

Solution

(a) Obtain ${}_{B_1}[I]_{B_2}$ using Lemma 13 (Topic 17D): i.e.

$${}_{B_1}[I]_{B_2} = ([\mathbf{w}_1]_{B_1}, [\mathbf{w}_2]_{B_1}).$$

We must then find a, b so that $\mathbf{w}_1 = a\mathbf{v}_1 + b\mathbf{v}_2$, that is we must solve:

$$\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}.$$

And we must also find c, d so that $\mathbf{w}_2 = c\mathbf{v}_1 + d\mathbf{v}_2$, that is we must solve:

$$\begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} 7 \\ 8 \end{pmatrix}.$$

Since these systems have the same coefficient matrix, we can solve them simultaneously using a super-augmented matrix:

$$\left(\begin{array}{cc|cc} 2 & -3 & 5 & 7 \\ -3 & 2 & 6 & 8 \end{array}\right) \longrightarrow \left(\begin{array}{cc|cc} 1 & \frac{-3}{2} & \frac{5}{2} & \frac{7}{2} \\ -3 & 2 & 6 & 8 \end{array}\right) \longrightarrow \left(\begin{array}{cc|cc} 1 & \frac{-3}{2} & \frac{5}{2} & \frac{7}{2} \\ 0 & \frac{-5}{2} & \frac{27}{2} & \frac{37}{2} \end{array}\right) \longrightarrow$$

$$\left(\begin{array}{cc|cc} 1 & \frac{-3}{2} & \frac{5}{2} & \frac{7}{2} \\ 0 & 1 & \frac{-27}{5} & \frac{-37}{5} \end{array}\right) \longrightarrow \left(\begin{array}{cc|cc} 1 & 0 & \frac{-28}{5} & \frac{-38}{5} \\ 0 & 1 & \frac{-27}{5} & \frac{-37}{5} \end{array}\right).$$

We see that:

$$a = \frac{-28}{5} \quad \text{and} \quad b = \frac{-27}{5}, \quad \text{so that} \quad \mathbf{w}_1 = \frac{-28}{5}\mathbf{v}_1 + \frac{-27}{5}\mathbf{v}_2$$

and

$$c = \frac{-38}{5} \quad \text{and} \quad d = \frac{-37}{5}, \quad \text{so that} \quad \mathbf{w}_2 = \frac{-38}{5}\mathbf{v}_1 + \frac{-37}{5}\mathbf{v}_2.$$

We thus obtain:

$${}_{B_1}[I]_{B_2} = \begin{pmatrix} \frac{-28}{5} & \frac{-38}{5} \\ \frac{-27}{5} & \frac{-37}{5} \end{pmatrix} = \frac{1}{5}\begin{pmatrix} -28 & -38 \\ -27 & -37 \end{pmatrix}.$$

Our second way of obtaining ${}_{B_1}[I]_{B_2}$ is to make use of the standard basis, and use the expression:

$${}_{B_1}[I]_{B_2} = {}_{B_1}[I]_S \; {}_S[I]_{B_2} \, ,$$

that is, to change coordinates from B_2 to B_1, we can go from B_2 to S and then from S to B_1.

The advantage of this approach is that ${}_S[I]_{B_i}$, for $i = 1, 2$, can both be written down immediately, and that we know ${}_{B_1}[I]_S = ({}_S[I]_{B_1})^{-1}$. We thus have:

$$
{}_S[I]_{B_2} = \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} \quad \text{and} \quad {}_{B_1}[I]_S = \begin{pmatrix} 2 & -3 \\ -3 & 2 \end{pmatrix}^{-1} = \frac{-1}{5} \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}, \qquad \text{so that}
$$

$$
{}_{B_1}[I]_{B_2} = {}_{B_1}[I]_S \; {}_S[I]_{B_2} = \frac{-1}{5} \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} = \frac{-1}{5} \begin{pmatrix} 28 & 38 \\ 27 & 37 \end{pmatrix}.
$$

(b) Using Corollary 3 (Topic 17D), we can write that:

$$
\begin{aligned}
{}_{B_2}[I]_{B_1} = ({}_{B_1}[I]_{B_2})^{-1} &= \left(\frac{-1}{5} \begin{pmatrix} 28 & 38 \\ 27 & 37 \end{pmatrix} \right)^{-1} \\
&= (-5) \frac{1}{10} \begin{pmatrix} 37 & -38 \\ -27 & 28 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -37 & 38 \\ 27 & -28 \end{pmatrix}.
\end{aligned}
$$

(c) If $[\mathbf{z}]_{B_2} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$, then we have that:

$$
[\mathbf{z}]_{B_1} = {}_{B_1}[I]_{B_2} \;\; [\mathbf{z}]_{B_2} = \frac{-1}{5} \begin{pmatrix} 28 & 38 \\ 27 & 37 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix} = \frac{-3}{5} \begin{pmatrix} 104 \\ 101 \end{pmatrix}.
$$

Also, if $[\mathbf{p}]_{B_1} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$, then we have that:

$$
[\mathbf{p}]_{B_2} = {}_{B_2}[I]_{B_1} \;\; [\mathbf{p}]_{B_1} = \frac{1}{2} \begin{pmatrix} -37 & 38 \\ 27 & -28 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 39 \\ -29 \end{pmatrix}.
$$

Topic 18

Matrix Representation of a Linear Operator

Previously, we have converted a linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ into its matrix representation in the standard basis: $[T]_S$.

Notice that when we do this, we are implicitly using the standard basis in both the domain, $\mathbb{F}^n$, and in the codomain, $\mathbb{F}^m$.

How do we obtain $[T]_S$?

We applied T to the standard basis vectors, $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ and we then put the (coordinates of the) images, that are expressed in the standard basis in $\mathbb{F}^m$, as columns of a matrix $[T]_S$, i.e.

$$[T]_S = ([T(\mathbf{e}_1)]_S, [T(\mathbf{e}_2)]_S, \ldots, [T(\mathbf{e}_n)]_S).$$

Why is this useful?

It can be used to obtain $T(\mathbf{v})$, as

$$T(\mathbf{v}) = [T(\mathbf{v})]_S = [T]_S [\mathbf{v}]_S.$$

Definition 1: Linear operator

Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. We refer to T as a **linear operator** to mean that $m = n$, so that $T : \mathbb{F}^n \to \mathbb{F}^n$.

We often say that T is a linear operator on $\mathbb{F}^n$, to mean that T is a linear operator from $\mathbb{F}^n$ to $\mathbb{F}^n$.

A linear operator is a special case of a linear transformation in which the domain and codomain are identical.

In this case, the matrix representation of T, will be a square matrix.

In addition, we can choose to use the same basis in both the domain and the codomain, and we will do so unless we tell you otherwise.

In this lecture we will restrict our attention to linear operators.

We have already had lots of examples, and, in fact, all of our geometric examples in Topic 13B were examples of linear operators.

We now discuss the consequences of using a basis in $\mathbb{F}^n$ other than the standard basis.

Definition 2

Let T a linear operator on $\mathbb{F}^n$, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{F}^n$.

We define the **matrix representation of T with respect to** B to be the matrix $[T]_B$ constructed as follows:

$$[T]_B = \left([T(\mathbf{v}_1)]_B\,, [T(\mathbf{v}_2)]_B\,, \ldots, [T(\mathbf{v}_n)]_B\right).$$

That is, $[T]_B$ is constructed by finding the images, under T, of the vectors in the basis B, and writing the (components of these) images, in the basis B, as the columns of a matrix.

Lemma 1

Let T be a linear operator on $\mathbb{F}^n$, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{F}^n$.

If $\mathbf{v} \in \mathbb{F}^n$, then

$$[T(\mathbf{v})]_B = [T]_B\ [\mathbf{v}]_B\,.$$

Proof

Since B is a basis for $\mathbb{F}^n$, and $\mathbf{v} \in \mathbb{F}^n$, then $\exists\ c_1, c_2, \ldots, c_n \in \mathbb{F}$ such that

$$\mathbf{v} = c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n, \quad \text{so that}$$

$$T(\mathbf{v}) = T(c_1\,\mathbf{v}_1 + c_2\,\mathbf{v}_2 + \cdots + c_n\,\mathbf{v}_n), \quad \text{and by linearity}$$

$$T(\mathbf{v}) = c_1\,T(\mathbf{v}_1) + c_2\,T(\mathbf{v}_2) + \cdots + c_n\,T(\mathbf{v}_n).$$

If we now take the coordinates, which is a linear operation, of this expression, we get :

$$\begin{aligned}
[T(\mathbf{v})]_B &= [c_1\,T(\mathbf{v}_1) + c_2\,T(\mathbf{v}_2) + \cdots + c_n\,T(\mathbf{v}_n)]_B \\
&= c_1\ [T(\mathbf{v}_1)]_B + c_2\ [T(\mathbf{v}_2)]_B + \cdots + c_n\ [T(\mathbf{v}_n)]_B \\
&= \left([T(\mathbf{v}_1)]_B\,, [T(\mathbf{v}_2)]_B\,, \ldots, [T(\mathbf{v}_n)]_B\right) \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = [T]_B\,[\mathbf{v}]_B\,. \quad \blacksquare
\end{aligned}$$

There are two important reasons why you may want to use another basis.

One is that you may have some geometrically or physically preferred vectors which naturally arise in your problem, such as the axis of rotation of an object.

Another reason is that you may wish to simplify the matrix representation, for example, it would be nice if $[T]_B$ were diagonal.

Often these two reasons are connected.

We will now revisit some of the geometric linear transformations from Topic 13B, and we will obtain their matrix representations in a basis which is adapted to each transformation. The simplicity of the resulting matrix representation should amaze you.

Example 1

Consider the linear operator, the projection onto the line $y = mx$ in $\mathbb{R}^2$.

We will refer to this function as T. Find a basis B such that $[T]_B$ is simple.

Solution

Let us think about some vectors which have very simple images.

The vector $\begin{pmatrix} 1 \\ m \end{pmatrix}$ lies on the line $y = mx$, and so its image is itself.

The vector $\begin{pmatrix} -m \\ 1 \end{pmatrix}$ lies orthogonal to the line, and so its image is $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

Let $B = \left\{ \begin{pmatrix} 1 \\ m \end{pmatrix}, \begin{pmatrix} -m \\ 1 \end{pmatrix} \right\}$. This set has two vectors that are not multiples of each other.

Thus it is a basis of $\mathbb{R}^2$.

Notice that

$$T\left(\begin{pmatrix} 1 \\ m \end{pmatrix}\right) = \begin{pmatrix} 1 \\ m \end{pmatrix} = 1\begin{pmatrix} 1 \\ m \end{pmatrix} + 0\begin{pmatrix} -m \\ 1 \end{pmatrix}, \text{ and}$$

$$T\left(\begin{pmatrix} -m \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} 1 \\ m \end{pmatrix} + 0\begin{pmatrix} -m \\ 1 \end{pmatrix}, \text{ and so}$$

$$[T]_B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

This is a very simple diagonal matrix, and it was obtained very quickly.

Example 2

Consider the linear transformation, the reflection across the plane $3x - 4y + 5z = 0$ in $\mathbb{R}^3$. We refer to this operator as L. Find a basis B such that $[L]_B$ is simple.

Solution

Once again, we consider some special vectors whose images are already known.

The vectors $\begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix}$, lie in the plane, and so their images are themselves.

The image of a normal vector however, such as $\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}$, is the negative of itself.

Let $B = \left\{ \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} \right\}$, then this is a set of three vectors in $\mathbb{R}^3$, and a quick check (how?) reveals that it is a linearly independent set, and thus a basis for $\mathbb{R}^3$.

We then have:

$$L\left(\begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix} = 1\begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix} + 0\begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix} + 0\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \text{ and}$$

$$L\left(\begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix}\right) = \begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix} = 0\begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix} + 0\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \text{ and}$$

$$L\left(\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}\right) = \begin{pmatrix} -3 \\ 4 \\ -5 \end{pmatrix} = 0\begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix} + 0\begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix} - 1\begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix}, \text{ and so}$$

$$[L]_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

This is a very simple diagonal matrix, and it was obtained very quickly.

You may not be satisfied with our results because you might prefer to use the standard basis instead, i.e. you may wish $[T]_S$ instead of $[T]_B$.

It should come as no surprise that the two matrices $[T]_B$ and $[T]_S$ are related. In fact, they are similar to each other.

Lemma 2

Let T be a linear operator on $\mathbb{F}^n$.

If B_1 and B_2 be bases for $\mathbb{F}^n$, then $[T]_{B_1}$ and $[T]_{B_2}$ are similar to each other, and

$$[T]_{B_2} = {}_{B_2}[I]_{B_1}\ [T]_{B_1}\ {}_{B_1}[I]_{B_2} = \left({}_{B_1}[I]_{B_2}\right)^{-1}\ [T]_{B_1}\ {}_{B_1}[I]_{B_2},$$

and

$$[T]_{B_1} = {}_{B_1}[I]_{B_2}\ [T]_{B_2}\ {}_{B_2}[I]_{B_1} = \left({}_{B_2}[I]_{B_1}\ \right)^{-1}\ [T]_{B_2}\ {}_{B_2}[I]_{B_1}.$$

Proof

Let $B_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and let $B_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$. Then

$$[T]_{B_2} = \left([T(\mathbf{w}_1)]_{B_2}, [T(\mathbf{w}_2)]_{B_2}, \ldots, [T(\mathbf{w}_n)]_{B_2}\right),$$

$$= \left({}_{B_2}[I]_{B_1}\,[T(\mathbf{w}_1)]_{B_1}, {}_{B_2}[I]_{B_1}\,[T(\mathbf{w}_2)]_{B_1}, \ldots, {}_{B_2}[I]_{B_1}\,[T(\mathbf{w}_n)]_{B_1}\right),$$

using the definition of matrix multiplication, we can write this as

$$[T]_{B_2} = {}_{B_2}[I]_{B_1}\left([T(\mathbf{w}_1)]_{B_1}, [T(\mathbf{w}_2)]_{B_1}, \ldots, [T(\mathbf{w}_n)]_{B_1}\right).$$

Recall that $[T(\mathbf{v})]_{B_1} = [T]_{B_1}\,[\mathbf{v}]_{B_1}$, and so we have:

$$[T]_{B_2} = {}_{B_2}[I]_{B_1}\ \left([T]_{B_1}\,[\mathbf{w}_1]_{B_1}, [T]_{B_1}\,[\mathbf{w}_2]_{B_1}, \ldots, [T]_{B_1}\,[\mathbf{w}_n]_{B_1}\right),$$

using the definition of matrix multiplication, we can write this as

$$[T]_{B_2} = {}_{B_2}[I]_{B_1}\ [T]_{B_1}\left([\mathbf{w}_1]_{B_1}, [\mathbf{w}_2]_{B_1}, \ldots, [\mathbf{w}_n]_{B_1}\right).$$

The proof is completed by noticing that this third matrix is ${}_{B_1}[I]_{B_2}$ and that ${}_{B_2}[I]_{B_1} = \left({}_{B_1}[I]_{B_2}\right)^{-1}$. Thus,

$$[T]_{B_2} = {}_{B_2}[I]_{B_1}\ [T]_{B_1}\ {}_{B_1}[I]_{B_2} = \left({}_{B_1}[I]_{B_2}\right)^{-1}\ [T]_{B_1}\ {}_{B_1}[I]_{B_2}.$$

To prove that $[T]_{B_1} = {}_{B_1}[I]_{B_2}\ [T]_{B_2}\ {}_{B_2}[I]_{B_1} = \left({}_{B_2}[I]_{B_1}\ \right)^{-1}\ [T]_{B_2}\ {}_{B_2}[I]_{B_1}$ is identical to the above proof, with the labels B_1 and B_2 interchanged. ■

Remark 1

If you read out the statement of Lemma 2, then the result should be logical and natural, it is very similar to the chain rule in calculus.

If you want the matrix representation of T in the basis B_2, then either work in basis B_2 entirely (the LHS), OR, change basis to B_1 then use the matrix representation of T in the basis B_1 and then finally change basis back to B_2.

Corollary 1

Let T be a linear operator on $\mathbb{F}^n$.

If B_1 be a basis for $\mathbb{F}^n$ and S the standard basis, then $[T]_{B_1}$ and $[T]_S$ are similar to each other, and

$$[T]_S = {}_S[I]_{B_1}\ [T]_{B_1}\ {}_{B_1}[I]_S = \left(({}_{B_1}[I]_S)^{-1}\ [T]_{B_1}\ {}_{B_1}[I]_S\right),$$

and

$$[T]_{B_1} = {}_{B_1}[I]_S\ [T]_S\ {}_S[I]_{B_1} = \left(({}_S[I]_{B_1}\ )^{-1}\ [T]_S\ {}_S[I]_{B_1}\right).$$

Proof

This is just a special case of Lemma 2 with B_2 replaced by S. ■

Example 3: Example 1 revisited.

Consider the linear transformation, the projection onto the line $y = mx$ in $\mathbb{R}^2$.

We will refer to this operator as T.

Let us make use of our solution to Example 1, and Corollary 1 to obtain $[T]_S$.

Solution

We derived, in Example 1, $[T]_B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, with $B = \left\{ \begin{pmatrix} 1 \\ m \end{pmatrix}, \begin{pmatrix} -m \\ 1 \end{pmatrix} \right\}$.

Thus we have :

$${}_S[I]_B = \begin{pmatrix} 1 & -m \\ m & 1 \end{pmatrix} \quad \text{and} \quad {}_B[I]_S = \frac{1}{1+m^2} \begin{pmatrix} 1 & m \\ -m & 1 \end{pmatrix}, \quad \text{and so}$$

$$[T]_S = {}_S[I]_B\ [T]_B\ {}_B[I]_S = \begin{pmatrix} 1 & -m \\ m & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{1+m^2} \begin{pmatrix} 1 & m \\ -m & 1 \end{pmatrix}$$

$$= \frac{1}{1+m^2} \begin{pmatrix} 1 & 0 \\ m & 0 \end{pmatrix} \begin{pmatrix} 1 & m \\ -m & 1 \end{pmatrix} = \frac{1}{1+m^2} \begin{pmatrix} 1 & m \\ m & m^2 \end{pmatrix},$$

as we found in Topic 13B, using the definition.

Example 4

Consider the linear transformation, the reflection across the plane $3x - 4y + 5z = 0$ in $\mathbb{R}^3$.

We refer to this operator as L.

Let us make us of our solution to Example 2 and Corollary 1 to obtain $[L]_S$.

We derived, in Example 2:

$$[L]_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \text{with} \;\; B = \left\{ \begin{pmatrix} 4 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \\ -3 \end{pmatrix} \begin{pmatrix} 3 \\ -4 \\ 5 \end{pmatrix} \right\}$$

Thus we have :

$${}_S[I]_B = \begin{pmatrix} 4 & 5 & 3 \\ 3 & 0 & -4 \\ 0 & -3 & 5 \end{pmatrix} \quad \text{and} \quad {}_B[I]_S = \frac{1}{150} \begin{pmatrix} 12 & 34 & 20 \\ 15 & -20 & -25 \\ 9 & -12 & 15 \end{pmatrix},$$

and so

$$\begin{aligned} [L]_S &= {}_S[I]_B \; [L]_B \; {}_B[I]_S \\ &= \begin{pmatrix} 4 & 5 & 3 \\ 3 & 0 & -4 \\ 0 & -3 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \frac{1}{150} \begin{pmatrix} 12 & 34 & 20 \\ 15 & -20 & -25 \\ 9 & -12 & 15 \end{pmatrix} \\ &= \frac{1}{150} \begin{pmatrix} 4 & 5 & 3 \\ 3 & 0 & -4 \\ 0 & -3 & 5 \end{pmatrix} \begin{pmatrix} 12 & 34 & 20 \\ 15 & -20 & -25 \\ -9 & 12 & -15 \end{pmatrix} \\ &= \frac{1}{150} \begin{pmatrix} 96 & 72 & -90 \\ 72 & 54 & 120 \\ -90 & 120 & 0 \end{pmatrix} = \frac{1}{25} \begin{pmatrix} 16 & 12 & -15 \\ 12 & 9 & 20 \\ -15 & 20 & 0 \end{pmatrix}. \end{aligned}$$

Topic 19A

Diagonalization of Linear Operators - I

Definition 1: Eigenvector, eigenvalue and eigenpair

Let $T : \mathbb{F}^n \to \mathbb{F}^n$, be a linear operator, then we say that the **non-zero** vector $\mathbf{x}$ is an **eigenvector** of T to mean that there exists a scalar $\lambda \in \mathbb{F}$ such that

$$T(\mathbf{x}) = \lambda\mathbf{x}.$$

This equation is called the **eigenvalue equation**, or the **eigenvalue problem**. The scalar λ is called an **eigenvalue** and the pair $(\lambda, \mathbf{x})$ is called an **eigenpair** of T.

Lemma 1

Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator. If B be a basis of $\mathbb{F}^n$, then

$(\lambda, \mathbf{x})$ is an **eigenpair** of T **iff** $(\lambda, [\mathbf{x}]_B)$ is an **eigenpair** of the matrix $[T]_B$.

Note that if S is the standard basis, then we would normally just write this statement as:

$(\lambda, \mathbf{x})$ is an **eigenpair** of T **iff** $(\lambda, \mathbf{x})$ is an **eigenpair** of the matrix $[T]_S$.

Proof

We just take the coordinates of both sides of the eigenvalue problem.

$$T(\mathbf{x}) = \lambda\mathbf{x} \Leftrightarrow [T(\mathbf{x})]_B = [\lambda\mathbf{x}]_B \Leftrightarrow [T]_B\,[\mathbf{x}]_B = \lambda\,[\mathbf{x}]_B\,. \qquad \blacksquare$$

Thus, solving the eigenvalue problem for a linear operator is as straightforward as solving the eigenvalue problem for a matrix.

Example 1

Find the eigenvalues of the linear operator given by

$$T : \mathbb{R}^2 \to \mathbb{R}^2, \quad \text{projection onto the line } y = mx.$$

For each eigenvalue, find a corresponding eigenvector.

Solution

We choose to work in the standard basis.
We have found (Topic 13B, Example 11) that

$$[T]_S = \frac{1}{1+m^2}\begin{pmatrix} 1 & m \\ m & m^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{1+m^2} & \frac{m}{1+m^2} \\ \frac{m}{1+m^2} & \frac{m^2}{1+m^2} \end{pmatrix}.$$

The characteristic polynomial of $[T]_S$ is:

$$\Delta_{[T]_S}(t) = \det\left(\begin{pmatrix} \frac{1}{1+m^2} - t & \frac{m}{1+m^2} \\ \frac{m}{1+m^2} & \frac{m^2}{1+m^2} - t \end{pmatrix}\right) = t^2 - t.$$

Thus T has eigenvalues, $\lambda_1 = 1, \lambda_2 = 0$. Let us now find the corresponding eigenvectors.

For $\lambda_1 = 1$, we have to solve:

$$([T]_S - 1I)\,\mathbf{x} = \mathbf{0}, \text{ that is }, \begin{pmatrix} \frac{-m^2}{1+m^2} & \frac{m}{1+m^2} \\ \frac{m}{1+m^2} & \frac{-1}{1+m^2} \end{pmatrix}\mathbf{x} = \mathbf{0},$$

which has solution set:

$$\left\{ s\begin{pmatrix} 1 \\ m \end{pmatrix} : s \in \mathbb{R} \right\}, \quad \text{and an eigenpair is} \quad \left(1, \begin{pmatrix} 1 \\ m \end{pmatrix}\right).$$

For $\lambda_2 = 0$, we have to solve:

$$([T]_S - 0I)\,\mathbf{x} = \mathbf{0}, \text{ is } \begin{pmatrix} \frac{1}{1+m^2} & \frac{m}{1+m^2} \\ \frac{m}{1+m^2} & \frac{m^2}{1+m^2} \end{pmatrix}\mathbf{x} = \mathbf{0},$$

which has solution set:

$$\left\{ t\begin{pmatrix} -m \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}, \quad \text{and an eigenpair is} \quad \left(0, \begin{pmatrix} -m \\ 1 \end{pmatrix}\right).$$

Definition 2: Diagonalizable

Let $T : \mathbb{F}^n \to \mathbb{F}^n$, be a linear operator. We say that T is **diagonalizable** to mean that there exists a basis B of $\mathbb{F}^n$ such that $[T]_B$ is a diagonal matrix.

Lemma 2

Let $T : \mathbb{F}^n \to \mathbb{F}^n$, be a linear operator. Then T is diagonalizable **iff** there exists a basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of $\mathbb{F}^n$ consisting of eigenvectors of T.

Proof

Let T be diagonalizable. Then exists a basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of $\mathbb{F}^n$ with

$$[T]_B = D = diag(d_1, d_2, \ldots, d_n).$$

In this case:

$$[T(\mathbf{v}_i)]_B = [T]_B\,[\mathbf{v}_i]_B = diag(d_1, d_2, \ldots, d_n)\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = d_i \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = d_i\,[\mathbf{v}_i]_B\,,$$

so that $\mathbf{v}_i$ is an eigenvector of T with eigenvalue d_i, for each $i = 1, 2, \ldots, n$.

On the other hand, if there exists a basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of $\mathbb{F}^n$ consisting of eigenvectors of T, then we have $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$, for some scalar λ_i, and this is true for all $i = 1, 2, \ldots, n$.

In this case, then $[T]_B = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and since this is a diagonal matrix, then T is diagonalizable. ■

Example 2

Show that the linear operator T: $\mathbb{R}^2 \to \mathbb{R}^2$, projection onto the line $y = mx$, is diagonalizable and find a new basis, B, for $\mathbb{R}^2$, in which $[T]_B$ is diagonal.

Solution

In Example 1 we found two eigenpairs of T:

$$\left(1, \begin{pmatrix} 1 \\ m \end{pmatrix}\right) \text{ and } \left(0, \begin{pmatrix} -m \\ 1 \end{pmatrix}\right).$$

We notice also that the two eigenvectors are not multiples of each other, and are thus linearly independent.

We let $B = \left\{ \begin{pmatrix} 1 \\ m \end{pmatrix}, \begin{pmatrix} -m \\ 1 \end{pmatrix} \right\}$. This is a basis for $\mathbb{R}^2$ and $[T]_B = diag(1, 0)$.

Notice that we obtained the same result in Topic 18, where we obtained this basis by thinking about the geometry of the transformation. That method will only be practical in some very special situations.

Lemma 3

Let $T : \mathbb{F}^n \to \mathbb{F}^n$, be a linear operator and let B_1 be a basis for $\mathbb{F}^n$. Then

$$T \text{ is diagonalizable } \textbf{ iff } \text{ the matrix } [T]_{B_1} \text{ is diagonalizable.}$$

Proof

We already know from Lemma 2 that T is diagonalizable **iff** there exists a basis B of $\mathbb{F}^n$ consisting of eigenvectors of T. In this case,

$$[T]_B = diag(\lambda_1, \lambda_2, \ldots, \lambda_n), \text{ with the } \lambda_i\text{'s being the eigenvalues of } T.$$

We also know from Topic 18, Lemma 2, that

$$[T]_B = {}_B[I]_{B_1}\ [T]_{B_1}\ {}_{B_1}[I]_B = \left({}_{B_1}[I]_B\right)^{-1}\ [T]_{B_1}\ {}_{B_1}[I]_B,$$

that is, the two matrices, $[T]_B$ and $[T]_{B_1}$, are similar, and the matrix relating them in the similarity transformation, is the change of basis matrix.

Thus, if T is diagonalizable, then there exists a basis B of $\mathbb{F}^n$ with

$$[T]_B = diag(\lambda_1, \lambda_2, \ldots, \lambda_n) = {}_B\,[I]_{B_1}\ [T]_{B_1}\ {}_{B_1}[I]_B,$$

so that $[T]_{B_1}$ is diagonalizable.

On the other hand, if $[T]_{B_1}$ is diagonalizable, then there exists an invertible matrix P, such that

$$P^{-1}\,[T]_{B_1}\,P = D = diag(d_1, d_2, \ldots, d_n).$$

We define the vectors $\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_n}$ by $[\mathbf{v}_1]_{B_1} = \mathbf{p}_1$, the first column of P, $[\mathbf{v}_2]_{B_1} = \mathbf{p}_2$, the second column of $P, \cdots, [\mathbf{v}_n]_{B_1} = \mathbf{p}_n$, the n^{th} column of P, so that

$$P = ([\mathbf{v}_1]_{B_1}, [\mathbf{v}_2]_{B_1}, \ldots, [\mathbf{v}_n]_{B_1}).$$

It follows that $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{F}^n$ since it contains n vectors in $\mathbb{F}^n$, and these are linearly independent because P is invertible and thus has rank of n. Moreover,

$$[T]_B = {}_B[I]_{B_1}\ [T]_{B_1}\ {}_{B_1}[I]_B = P^{-1}\,[T]_{B_1}\,P = D. \qquad \blacksquare$$

Consequences

I) To determine whether a linear operator T is diagonalizable or not, we can obtain the matrix representation of T in any basis B_1, $[T]_{B_1}$, and test whether this matrix is diagonalizable.

II) If $[T]_{B_1}$ is diagonalizable and $P^{-1}\,[T]_{B_1}\,P = D$ is a diagonal matrix, then

(a) the entries of D are the eigenvalues of the linear operator T, and

(b) the columns of the matrix P are the components $[\mathbf{v}_i]_{B_1}$ of the eigenvectors of T in the basis B_1. In particular, if you use the basis $B_1 = S$, then the columns of P are the components $[\mathbf{v}_i]_S = \mathbf{v}_i$ of the eigenvectors of T in the basis S.

Example 3

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$, be a linear operator and let $[T]_S = \begin{pmatrix} 3 & 4 & -2 \\ 3 & 8 & -3 \\ 6 & 14 & -5 \end{pmatrix} = A.$

Show that T is diagonalizable and find a new basis B for $\mathbb{R}^3$ in which $[T]_B$ is a diagonal matrix.

Solution

The arithmetical part of this solution is one that we have done many times before, that is, we will diagonalize this matrix A by determining its eigenpairs.

$$\Delta_A(t) = \det \begin{pmatrix} 3-t & 4 & -2 \\ 3 & 8-t & -3 \\ 6 & 14 & -5-t \end{pmatrix} = -t^3 + 6t^2 - 11t + 6 = -(t-3)(t-2)(t-1).$$

The eigenvalues of A are $\lambda_1 = 3, \lambda_2 = 2, \lambda_3 = 1$. Let us find the corresponding eigenvectors.

i) For $\lambda_1 = 3$, we examine

$$(A - 3I)\mathbf{v} = \mathbf{0}\,, \quad \text{which is equivalent to} \begin{pmatrix} 0 & 4 & -2 \\ 3 & 5 & -3 \\ 6 & 14 & -8 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{which row reduces to} \begin{pmatrix} 6 & 0 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

A non-trivial solution is $\mathbf{v} = (1, 3, 6)^T$.

ii) For $\lambda_2 = 2$, we examine

$$(A-2I)\mathbf{v}=\mathbf{0}\,,\quad \text{which is equivalent to } \begin{pmatrix} 1 & 4 & -2 \\ 3 & 6 & -3 \\ 6 & 14 & -7 \end{pmatrix}\mathbf{v}=\mathbf{0},$$

$$\text{which row reduces to } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{pmatrix}\mathbf{v}=\mathbf{0}.$$

A non-trivial solution is $\mathbf{v}=(0,1,2)^T$.

iii) For $\lambda_3 = 1$, we examine

$$(A-1I)\mathbf{v}=\mathbf{0},\quad \text{which is equivalent to } \begin{pmatrix} 2 & 4 & -2 \\ 3 & 7 & -3 \\ 6 & 14 & -6 \end{pmatrix}\mathbf{v}=\mathbf{0},$$

$$\text{which row reduces to } \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\mathbf{v}=\mathbf{0}.$$

A non-trivial solution is $\mathbf{v}=(1,0,1)^T$.

The matrix A is diagonalizable, and if we let $P=\begin{pmatrix} 1 & 0 & 1 \\ 3 & 1 & 0 \\ 6 & 2 & 1 \end{pmatrix}$, then we have $P^{-1}AP = D = diag(3,2,1)$.

We also conclude that if we let $B=\left\{\begin{pmatrix} 1 \\ 3 \\ 6 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\right\}$, then $[T]_B = diag(3,2,1)$.

Corollary 1

Let $A \in M_{n\times n}(\mathbb{F})$. Then

A is diagonalizable **iff** there exists a basis of $\mathbb{F}^n$ of eigenvectors of A.

Proof

Let us introduce the linear operator $T_A : \mathbb{F}^n \to \mathbb{F}^n$, such that $T_A(\mathbf{x}) = A\mathbf{x}$, $[T]_S = A$.

Lemma 3 tells us that T_A is diagonalizable **iff** $[T]_S = A$ is diagonalizable.

Lemma 2 tells us that T_A is diagonalizable **iff** there exists a basis B of $\mathbb{F}^n$ consisting of eigenvectors of T_A.

Lemma 1 tells us that $\mathbf{x}$ is an eigenvector of T_A **iff** $\mathbf{x}$ is an eigenvector of A.

Thus A is diagonalizable **iff** there exists a basis B of $\mathbb{F}^n$ consisting of eigenvectors of A. ■

As a consequence of Lemma 3, we need only consider the issue of diagonalization of n by n matrices in the future.

In addition, Corollary 1 tells us that we need to examine these matrices for the existence of n linearly independent eigenvectors. This is the issue which we will address next.

Lemma 4

Let $A \in M_{n\times n}(\mathbb{F})$ have eigenpairs $(\lambda_1, \mathbf{v}_1), (\lambda_2, \mathbf{v}_2), \ldots, (\lambda_m, \mathbf{v}_m)$, for $1 \le m \le n$.

If the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ are all different, then the set $W = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is linearly independent.

Proof

We prove the result by contradiction.

Let us assume that $W = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is linearly dependent and let us reduce this set to a linearly independent set $V = \{\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}\}$ which has the same span as W, using Lemma 5 of Topic 17B.

The proof of Lemma 5 in Topic 17B also shows us that if $\mathbf{v}_k \in W$ and $\mathbf{v}_k \notin V$, then $\mathbf{v}_k$ can be written as a linear combination of the vectors in V, that is,

$$\begin{aligned}
(i) \qquad \mathbf{v}_k &= c_1\mathbf{v}_{q_1} + c_2\mathbf{v}_{q_2} + \cdots + c_r\mathbf{v}_{q_r}, && \text{applying } A \text{ gives}\\
A\mathbf{v}_k &= A(c_1\mathbf{v}_{q_1} + c_2\mathbf{v}_{q_2} + \cdots + c_r\mathbf{v}_{q_r}), && \text{using linearity yields}\\
A\mathbf{v}_k &= c_1A\mathbf{v}_{q_1} + c_2A\mathbf{v}_{q_2} + \cdots + c_rA\mathbf{v}_{q_r}, && \text{since all } \mathbf{v}_j\text{'s are eigenvectors, we get}\\
(ii) \quad \lambda_k\mathbf{v}_k &= c_1\lambda_{q_1}\mathbf{v}_{q_1} + c_2\lambda_{q_2}\mathbf{v}_{q_2} + \cdots + c_r\lambda_{q_r}\mathbf{v}_{q_r}\,.
\end{aligned}$$

If we now form λ_k times (i) and subtract (ii), we get

$$\mathbf{0} = c_1(\lambda_k - \lambda_{q_1})\mathbf{v}_{q_1} + c_2(\lambda_k - \lambda_{q_2})\mathbf{v}_{q_2} + \cdots + c_r(\lambda_k - \lambda_{q_r})\mathbf{v}_{q_r}.$$

Since the set $V = \{\mathbf{v}_{q_1}, \mathbf{v}_{q_2}, \ldots, \mathbf{v}_{q_r}\}$ is linearly independent, then all the coefficients in this expression must be zero, that is:

$$0 = c_1(\lambda_k - \lambda_{q_1}) = c_2(\lambda_k - \lambda_{q_2}) = \cdots = c_r(\lambda_k - \lambda_{q_r}).$$

And since the eigenvalues are all distinct, we have

$$(\lambda_k - \lambda_{q_1}) \neq 0\,, (\lambda_k - \lambda_{q_2}) \neq 0\,, \ldots, (\lambda_k - \lambda_{q_r}) \neq 0\,,$$

we conclude that all the c_i's are zero.

Due to expression (i), we now have the contradiction that the eigenvector $\mathbf{v}_k$ is the zero vector. We conclude that $W = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ must be linearly independent. $\blacksquare$

An immediate consequence of this result is Lemma 2 in Topic 16B, that is,

> Let $A \in M_{n\times n}(\mathbb{F})$ have eigenpairs $(\lambda_1, \mathbf{v}_1), (\lambda_2, \mathbf{v}_2), \ldots, (\lambda_n, \mathbf{v}_n)$, where the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all different.
> If $P = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$, then P is invertible and $P^{-1}AP = D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

Proof: If we have n eigenvector corresponding to n different eigenvalues, then these n vectors are linearly independent by Lemma 4, and thus form a basis for $\mathbb{F}^n$, and this is a basis of eigenvectors of the matrix A.
Thus A is diagonalizable by Corollary 1 of Lemma 3, and so it is similar to a diagonal matrix.

If we let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, then $[T_A]_B = D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, while $[T_A]_S = A$, and ${}_S[I]_B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = P$, and so

$$P^{-1}AP = \left({}_S[I]_B\right)^{-1} \ [T_A]_S \ {}_S[I]_B = [T_A]_B = D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n). \qquad \blacksquare$$

Definition 3: Characteristic polynomial

Let $T : \mathbb{F}^n \to \mathbb{F}^n$, be a linear operator, and let B be a basis for $\mathbb{F}^n$.

We define the **characteristic polynomial** of T, $\Delta_T(t)$, to mean

$$\Delta_T(t) = \Delta_{[T]_B}(t).$$

Note that this quantity would only be meaningful if it did not depend upon the basis B that is used.

Example 4

Suppose $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear operator with $[T]_S = \begin{pmatrix} 5 & 2 & -1 \\ 8 & 1 & -2 \\ 16 & 0 & -3 \end{pmatrix}$.

Evaluate $\Delta_T(t)$ and use it to prove that T is diagonalizable.

Find a basis, B, of $\mathbb{R}^3$ such that $[T]_B$ is a diagonal matrix.

Solution

The characteristic polynomial of T is:

$$\Delta_T(t) = \det\left(\begin{pmatrix} 5-t & 2 & -1 \\ 8 & 1-t & -2 \\ 16 & 0 & -3-t \end{pmatrix}\right) = -t^3 + 3t^2 + 13t - 15.$$

We note that $-t^3+3t^2+13t-15 = -(t-5)(t-1)(t+3)$, so that the eigenvalues of T are $\lambda_1 = 5, \lambda_2 = 1, \lambda_3 = -3$. We know now that T is diagonalizable and thus need to obtain the corresponding eigenvectors.

i) for $\lambda_1 = 5$, we examine

$$(A-5I)\mathbf{v} = \mathbf{0}, \quad \text{which is equivalent to} \begin{pmatrix} 0 & 2 & -1 \\ 8 & -4 & -2 \\ 16 & 0 & -8 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{which row reduces to} \begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

And a non-trivial solution is $\mathbf{v} = (1,1,2)^T$.

ii) For $\lambda_2 = 1$, we examine

$$(A-(1)I)\mathbf{v} = \mathbf{0}, \quad \text{which is equivalent to} \begin{pmatrix} 4 & 2 & -1 \\ 8 & 0 & -2 \\ 16 & 0 & -4 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{which row reduces to} \begin{pmatrix} 4 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

And a non-trivial solution is $\mathbf{v} = (1,0,4)^T$.

iii) For $\lambda_3 = -3$, we examine

$$(A+3I)\mathbf{v} = \mathbf{0}, \quad \text{which is equivalent to} \begin{pmatrix} 8 & 2 & -1 \\ 8 & 4 & -2 \\ 16 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{which row reduces to} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

And a non-trivial solution is $\mathbf{v} = (0,1,2)^T$.

We conclude that:

$$B = \left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\},$$

$$[T]_B = diag(5,1,-3),$$

and

$$P^{-1}AP = diag(5,1,-3), \quad \text{with} \quad P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 4 & 2 \end{pmatrix}.$$

Topic 19B

Diagonalization of Linear Operators - II

The main goal of this lecture is to determine conditions on a matrix $A \in M_{n\times n}(\mathbb{F})$ for it to possess n linearly independent eigenvectors, so that it will be diagonalizable.

There are two possible areas of concern.

The first problem arises when there less than n eigenvalues: this can only happens when $\mathbb{F} = \mathbb{R}$.

Example 5

Let $A \in M_{2\times 2}(\mathbb{R})$ and$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, then $\Delta_A(t) = t^2 + 1$.

This matrix has no (real) eigenvalues, and no (real) eigenvectors: it is not diagonalizable.

Example 6

Let $B \in M_{3\times 3}(\mathbb{R})$ and $B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$, then $\Delta_B(t) = (2-t)(t^2+1)$.

This matrix has only one (real) eigenvalue, $\lambda_1 = 2$, and only one linearly independent eigenvector, $(1,0,0)^T$, for example (or any non-zero multiple of this), and so it is not diagonalizable.

The second issue is when there are less than n linearly independent eigenvectors, even though there are n eigenvalues.

Example 7

Let $C \in M_{2\times 2}(\mathbb{R})$ and

$$C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \text{then } \Delta_C(t) = t^2 = (t-0)^2.$$

This (2×2) matrix has 2 eigenvalues, they are both zero. However, the solution set to

$$(C - 0I)\,\mathbf{x} = \mathbf{0} \;\text{ is }\; Span\left(\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}\right),$$

thus we can obtain only one linearly independent eigenvector for C.

We will now introduce some theory to explain these deficiencies.

Definition 4: Algebraic multiplicity

Let λ be an eigenvalue of $A \in M_{n\times n}(\mathbb{F})$. We say that the **algebraic multiplicity** of λ is a_λ to mean that a_λ is the highest power of the factor $(t-\lambda)$ that divides the characteristic polynomial $\Delta_A(t)$.

That is, $(t-\lambda)^{a_\lambda} | \Delta_A(t)$, but $(t-\lambda)^{a_\lambda+1} \nmid \Delta_A(t)$.

Example 8

Let $B \in M_{3\times 3}(\mathbb{R})$ and

$$B = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}, \quad \text{then } \Delta_B(t) = -(t-5)(t+1)^2.$$

The matrix B has eigenvalues of $\lambda_1 = 5$, with algebraic multiplicity $a_5 = 1$, and $\lambda_2 = -1$, with algebraic multiplicity $a_{-1} = 2$.

We think of this matrix B has having three eigenvalues, and one of them is repeated twice, so that its eigenvalues are $5, -1$, and -1. The second eigenvalue, $\lambda_2 = -1$, has an algebraic multiplicity of 2, and this tells us that we need to consider this eigenvalue as occurring twice.

The algebraic multiplicity of an eigenvalue tells us *how many linearly independent eigenvectors* ***we would like to obtain*** *from that eigenvalue*. In the simplest case, each eigenvalue has an algebraic multiplicity of 1, and contributes one linearly independent eigenvector.

Definition 5: Geometric multiplicity

Let λ be an eigenvalue of $A \in M_{n\times n}(\mathbb{F})$. We say that the **geometric multiplicity** of λ is g_λ to mean that g_λ is equal the dimension of the eigenspace E_λ.

The dimension of E_λ refers to the number of vectors in a basis of E_λ.

Example 9

Let $B \in M_{3\times 3}(\mathbb{R})$ and

$$B = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}, \quad \text{then } \Delta_B(t) = -(t-5)(t+1)^2.$$

The matrix B has an eigenvalue of $\lambda_1 = 5$, we obtain E_5 by solving:

$$(B - 5I)\mathbf{x} = \mathbf{0}, \quad \text{which is} \begin{pmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{pmatrix} \mathbf{x} = \mathbf{0},$$

$$\text{which reduces to} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0},$$

$$\text{with solution set, } E_5 = \mathit{Span}\left(\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}\right).$$

We notice that E_5 has one linearly independent vector in its spanning set and so $g_5 = 1$.

The matrix B has an eigenvalue of $\lambda_2 = -1$, we obtain E_{-1} by solving:

$$(B + 1I)\mathbf{x} = \mathbf{0}, \quad \text{which is} \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix} \mathbf{x} = \mathbf{0},$$

$$\text{which reduces to} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0},$$

$$\text{with solution set, } E_{-1} = \mathit{Span}\left(\left\{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\right\}\right).$$

We notice that E_{-1} has two linearly independent vectors in its spanning set and so $g_{-1} = 2$.

The geometric multiplicity of an eigenvalue tells us *how many linearly independent eigenvectors* ***we can obtain*** *from that eigenvalue.*

The two multiplicities are connected through the following lemma:

Lemma 5

Let λ be an eigenvalue of the matrix $A \in M_{n\times n}(\mathbb{F})$.

If the geometric multiplicity of λ is g_λ, and the algebraic multiplicity of λ is a_λ, then

$$1 \leq g_\lambda \leq a_\lambda.$$

Proof

If λ be is an eigenvalue of A, then there is a non-trivial solution of $A\mathbf{v} = \lambda\mathbf{v}$, so $1 \leq g_\lambda$.

Suppose that $B_1 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\}$ is a basis for E_λ, where $p = g_\lambda$.

We extend this basis to a basis B of $\mathbb{F}^n$, for example, by adding (suitable) $(n-p)$ vectors from the standard basis, i.e. $B = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p, \mathbf{u}_1, \ldots, \mathbf{u}_{n-p}\}$.

We note that $A\mathbf{w}_i = \lambda\mathbf{w}_i$, for $i = 1, \ldots, p$, since each $\mathbf{w}_i \in E_\lambda$. It now follows that $[T_A]_B$ has the following structure:

$$[T_A]_B = \left(\begin{array}{cc} \begin{array}{cccc} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{array} & M_1 \\ \mathbb{O}_{(n-p)\times p} & M_2 \end{array} \right),$$

for some $M_1 \in M_{p\times(n-p)}$ and some $M_2 \in M_{(n-p)\times(n-p)}$.

It may be shown by induction that

$$\Delta_A(t) = (\lambda - t)^p \Delta_{M_2}(t) = (\lambda - t)^{g_\lambda} \Delta_{M_2}(t).$$

Since a_λ is the largest power of $(t-\lambda)$ which divides $\Delta_A(t)$, it follows that $a_\lambda \geq g_\lambda$. ■

Given an eigenvalue, λ, we would like there to be a_λ linearly independent eigenvectors in E_λ. However, there are only g_λ linearly independent eigenvectors in E_λ. The next result tells us about how many linearly independent eigenvectors we will have altogether.

Lemma 6

Let $A \in M_{n\times n}(\mathbb{F})$ have different eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ and suppose that their corresponding eigenspaces, $E_{\lambda_1}, E_{\lambda_2}, \ldots, E_{\lambda_m}$ have bases $B_1, B_2, \ldots, B_m$. Then

$$B = B_1 \cup B_2 \cup \cdots \cup B_m \text{ is linearly independent.}$$

Proof

Let $B = B_1 \cup B_2 \cup \cdots \cup B_m$, $B_1 = \{\mathbf{v}_{11}, \mathbf{v}_{12}, \cdots, \mathbf{v}_{1g_{\lambda_1}}\}, \ldots, B_m = \{\mathbf{v}_{m1}, \mathbf{v}_{m2}, \cdots, \mathbf{v}_{mg_{\lambda_m}}\}$.

Let us consider a linear combination of vectors in B, so that the first g_{λ_1} are from B_1, the next g_{λ_2} are from B_2, ..., and the last g_{λ_m} are from B_m, and equate it to zero., that is, we examine

$$\sum_{i=1}^{i=g_{\lambda_1}} c_{1i}\mathbf{v}_{1i} + \sum_{i=1}^{i=g_{\lambda_2}} c_{2i}\mathbf{v}_{2i} + \cdots + \sum_{i=1}^{i=g_{\lambda_m}} c_{mi}\mathbf{v}_{mi} = \mathbf{0}\ (*)$$

The vector $\sum_{i=1}^{i=g_{\lambda_1}} c_{1i}\mathbf{v}_{1i} = \mathbf{v}_1$ is a vector in E_{λ_1}, and so it is either an eigenvector of A with eigenvalue λ_1, or the zero vector.

The vector $\sum_{i=1}^{i=g_{\lambda 2}} c_{2i}\mathbf{v}_{2i} = \mathbf{v}_2$ is a vector in E_{λ_2}, and so it is either an eigenvector of A with eigenvalue λ_2, or the zero vector.

$\vdots$

$\vdots$

The vector $\sum_{i=1}^{i=g_{\lambda_m}} c_{mi}\mathbf{v}_{mi} = \mathbf{v}_m$ is a vector in E_{λ_m}, and so it is either an eigenvector of A with eigenvalue λ_m, or the zero vector.

We then have

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

The vectors in this sum are either eigenvectors corresponding to different eigenvalues, or they are the zero vector. Since we know from Lemma 4 of Topic 19A, that eigenvectors corresponding to different eigenvalues are linearly independent, then we conclude that each of the vectors in this sum is the zero vector and thus that all of the constants appearing in equation $(*)$ are zero. Thus B is linearly independent. ■

We are now in a position to count the number of linearly independent eigenvectors that a matrix $A \in M_{n\times n}(\mathbb{F})$ possesses.

Lemma 7

Let $A \in M_{n\times n}(\mathbb{F})$ have $\Delta_A(t) = (\lambda_1 - t)^{a_{\lambda_1}}(\lambda_2 - t)^{a_{\lambda_2}} \ldots (\lambda_m - t)^{a_{\lambda_m}}\, h(t)$, where $\lambda_i,\ i = 1, \ldots, m$, are different eigenvalues of A, with corresponding algebraic multiplicities a_{λ_i}, and $h(t)$ is a polynomial in t with no linear factors. Then

A is diagonalizable **iff** both $h(t) = 1$ and $a_{\lambda_i} = g_{\lambda_i}$, for each $i = 1, 2, \ldots, m$.

Proof

We already know that A is diagonalizable **iff** A has n linearly independent eigenvectors. Let us count the linearly independent eigenvectors of A.

We consider the set $B = B_1 \cup B_2 \cup \cdots \cup B_m$ introduced in Lemma 6.

λ_1 contributes exactly g_{λ_1} linearly independent eigenvectors to B,

λ_2 contributes exactly g_{λ_2} linearly independent eigenvectors to B,

$\vdots$

λ_m contributes exactly g_{λ_m} linearly independent eigenvectors to B.

We thus have a total of $N = g_{\lambda_1} + g_{\lambda_2} + \cdots + g_{\lambda_m}$ linearly independent vectors in B, with

$$(i) \qquad N = \sum_{i=1}^{i=m} g_{\lambda_i} \le \sum_{i=1}^{i=m} a_{\lambda_i},$$

with equality **iff** $g_{\lambda_i} = a_{\lambda_i}$, for every $i = 1, 2, \ldots, m$ (using Lemma 5).

We also know that $\Delta_A(t)$ is an n^{th} order polynomial with coefficient of t^n of $(-1)^n$, so that

$$(ii) \qquad n = \sum_{i=1}^{i=m} a_{\lambda_i} + \text{the order of } h(t).$$

Combining expressions (i) and (ii) gives

$$N = \sum_{i=1}^{i=m} g_{\lambda_i} \le \sum_{i=1}^{i=m} a_{\lambda_i} = n - \text{the order of } h(t).$$

We see that the only way that $N = n$ is if the order of $h(t) = 0$, so that the polynomial $h(t) = 1$, and if $g_{\lambda_i} = a_{\lambda_i}$, for every $i = 1, 2, \ldots, m$. ■

Example 10

We return to Examples 8 and 9, in which we had the matrix $B = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$.

This matrix is diagonalizable. We know this because $\Delta_B(t) = -(t-5)(t+1)^2$.
This polynomial has only linear factors, and we have, $a_5 = g_5 = 1$, and $a_{-1} = g_{-1} = 2$.

Example 11

Is $C = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{pmatrix}$ diagonalizable?

Solution

We have $\Delta_C(t) = (t-3)^2(t+1)^2$. So that $\lambda_1 = 3$, with $a_{\lambda_1} = 2$, and $\lambda_2 = -1$, with $a_{\lambda_2} = 2$.

As for their eigenspaces, we must examine:

$$(C - 3I)\mathbf{v} = \mathbf{0}, \quad \text{which is} \begin{pmatrix} -2 & 2 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & 2 & -2 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{which reduces to} \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

Since the coefficient matrix has rank of 2, then there will be $4 - 2 = 2$ parameters in the solution set and thus E_3 is two dimensional. We also examine:

$$(C + 1I)\mathbf{v} = \mathbf{0}, \quad \text{which is} \begin{pmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{which reduces to} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

Since the coefficient matrix has rank of 2, then there will be $4 - 2 = 2$ parameters in the solution set and thus E_{-1} is two dimensional.

Since $a_3 = g_3 = 2$, and $a_{-1} = g_{-1} = 2$, we conclude that the matrix C is diagonalizable.

Example 12

Is $D = \begin{pmatrix} 5 & 2 & 0 & 1 \\ -2 & 1 & 0 & -1 \\ 4 & 4 & 3 & 2 \\ 16 & 0 & -8 & -5 \end{pmatrix}$ diagonalizable?

Solution

We have $\Delta_D(t) = (t-3)^3(t+5)^1$. So that $\lambda_1 = 3$, with $a_{\lambda_1} = 3$, and $\lambda_2 = -5$, with $a_{\lambda_2} = 1$.

As for their eigenspaces, we must examine:

$$(D - 3I)\mathbf{v} = \mathbf{0}, \quad \text{which is} \begin{pmatrix} 2 & 2 & 0 & 1 \\ -2 & -2 & 0 & -1 \\ 4 & 4 & 0 & 2 \\ 16 & 0 & -8 & -8 \end{pmatrix} \mathbf{v} = \mathbf{0}$$

$$\text{which reduces to} \begin{pmatrix} 2 & 2 & 0 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0}.$$

Since the coefficient matrix has rank of 2, then there will be $4 - 2 = 2$ parameters in the solution set and thus E_3 is two dimensional.

Since $3 = a_3 \neq g_3 = 2$, we conclude that the matrix D is not diagonalizable.

Topic 19C

Diagonalization of Linear Operators - III

Application of diagonalization tin obtaining powers of a matrix

Let $A \in M_{n\times n}(\mathbb{F})$ and suppose that A is diagonalizable, then we know that there exists an invertible matrix P, such that

$$P^{-1}AP = D = diag(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where λ_i , $i = 1, \dots, n$, are the eigenvalues of A (not necessarily all distinct).

We note that for any positive integer k, we have that:

$$D^k = diag(\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k),$$

and also that

$$\left(P^{-1}AP\right)^k = P^{-1}A^kP = D^k.$$

We thus conclude that :

$$A^k \text{ is diagonalizable and } A^k = PD^kP^{-1}.$$

Notice the RHS of this equation is relatively simple to evaluate as opposed to actually multiplying A by itself, k-times.

Example 13

Find a simple expression for A^k when $A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$.

Solution

We first examine the eigenvalue problem for the matrix A.

The characteristic polynomial is:

$$\Delta_A(t) = \det\begin{pmatrix} -t & -2 \\ 1 & 3-t \end{pmatrix} = (t-2)(t-1).$$

For the eigenvalue $\lambda_1 = 2$, we have:

$$(A - 2I)\mathbf{v} = \mathbf{0}, \quad \text{which is } \begin{pmatrix} -2 & -2 \\ 1 & 1 \end{pmatrix}\mathbf{v} = \mathbf{0},$$

$$\text{which reduces to } \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{and the eigenspace is } Span\left(\left\{(1,-1)^T\right\}\right).$$

For the eigenvalue $\lambda_2 = 1$, we have:

$$(A - 1I)\mathbf{v} = \mathbf{0}, \text{ which is } \begin{pmatrix} -1 & -2 \\ 1 & 2 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{which reduces to } \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

$$\text{and the eigenspace is } Span\left(\left\{(2,-1)^T\right\}\right).$$

We then conclude that A is diagonalizable:

$$P^{-1}AP = diag(2,1), \quad \text{where}$$

$$P = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix} \text{ and } P^{-1} = \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}.$$

It now follows that,

$$A^k = PD^kP^{-1} = \begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}^k \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2^k & 2 \\ -2^k & -1 \end{pmatrix} \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}$$

$$\text{and thus } A^k = \begin{pmatrix} 2-2^k & 2-2^{k+1} \\ 2^k-1 & 2^{k+1}-1 \end{pmatrix}.$$

Example 14

Use diagonalization to obtain an expression for A^{16}, if $A = \begin{pmatrix} -7 & -20 & -8 \\ 1 & 2 & 2 \\ 4 & 10 & 5 \end{pmatrix}$.

Solution

We first solve the eigenvalue problem for A :

$$\Delta_A(t) = \det \begin{pmatrix} -7-t & -20 & -8 \\ 1 & 2-t & 2 \\ 4 & 10 & 5-t \end{pmatrix} = -(t^3 - 7t + 6) = -(t-2)(t-1)(t+3).$$

For the eigenvalue $\lambda_1 = 2$, we examine: $(A - 2I)\mathbf{v} = \mathbf{0}$, which is:

$$\begin{pmatrix} -9 & -20 & -8 \\ 1 & 0 & 2 \\ 4 & 10 & 3 \end{pmatrix} \mathbf{v} = \mathbf{0}, \text{ reducing to } \begin{pmatrix} 1 & 0 & 2 \\ 0 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

and the eigenspace is $Span\left(\left\{(-4, 1, 2)^T\right\}\right)$.

For the eigenvalue $\lambda_2 = 1$, we examine: $(A - 1I)\mathbf{v} = \mathbf{0}$, which is:

$$\begin{pmatrix} -8 & -20 & -8 \\ 1 & 1 & 2 \\ 4 & 10 & 4 \end{pmatrix} \mathbf{v} = \mathbf{0}, \text{ reducing to } \begin{pmatrix} 1 & 0 & \frac{8}{3} \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

and the eigenspace is $Span\left(\left\{(-8, 2, 3)^T\right\}\right)$.

For the eigenvalue $\lambda_3 = -3$, we examine: $(A + 3I)\mathbf{v} = \mathbf{0}$, which is:

$$\begin{pmatrix} -4 & -20 & -8 \\ 1 & 5 & 2 \\ 4 & 10 & 8 \end{pmatrix} \mathbf{v} = \mathbf{0}, \text{ reducing to } \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

and the eigenspace is $Span\left(\left\{(-2, 0, 1)^T\right\}\right)$.

We have $P^{-1}AP = diag(2, 1, -3)$,

$$\text{with } P = \begin{pmatrix} -4 & -8 & -2 \\ 1 & 2 & 0 \\ 2 & 3 & 1 \end{pmatrix}, \text{ and so } P^{-1} = \frac{1}{2}\begin{pmatrix} 2 & 2 & 4 \\ -1 & 0 & -2 \\ -1 & -4 & 0 \end{pmatrix}.$$

It now follows that,

$$\begin{aligned} A^{16} &= PD^{16}P^{-1} \\ &= \begin{pmatrix} -4 & -8 & -2 \\ 1 & 2 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix}^{16} \frac{1}{2}\begin{pmatrix} 2 & 2 & 4 \\ -1 & 0 & -2 \\ -1 & -4 & 0 \end{pmatrix} \\ &= \begin{pmatrix} (-4)(2^{16}) & -8 & (-2)(-3)^{16} \\ (2^{16}) & 2 & 0 \\ 2(2^{16}) & 3 & (-3)^{16} \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ -\frac{1}{2} & 0 & -1 \\ -\frac{1}{2} & -2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} (-4)(2^{16}) + 4 + (-3)^{16} & (-4)(2^{16}) + 4(-3)^{16} & (-8)[(2^{16}) - 1] \\ (2^{16}) - 1 & (2^{16}) & 2[(2^{16}) - 1] \\ (-\frac{1}{2})\left[(-2)(2^{16}) + 3 + (-3)^{16}\right] & 2[(2^{16}) - (-3)^{16}] & 4(2^{16}) - 3 \end{pmatrix}. \end{aligned}$$

Example 15

Find a simple expression for B^k when $B = \begin{pmatrix} 0 & -4 \\ 1 & 2 \end{pmatrix}$.

Solution

We first examine the eigenvalue problem for the matrix B.

$$\Delta_B(t) = \det \begin{pmatrix} -t & -4 \\ 1 & 2-t \end{pmatrix} = t^2 - 2t + 4 = (t - 1 - ir)(t - 1 + ir), \quad \text{where } r = \sqrt{3}.$$

For the eigenvalue $\lambda_1 = 1 + ir = \alpha$, we examine: $(B - \alpha\, I)\mathbf{v} = \mathbf{0}$, which is:

$$\begin{pmatrix} -\alpha & -4 \\ 1 & \overline{\alpha} \end{pmatrix} \mathbf{v} = \mathbf{0}, \text{ reducing to } \begin{pmatrix} 1 & \overline{\alpha} \\ 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

and the eigenspace is $Span\left(\left\{(\overline{\alpha}, -1)^T\right\}\right)$.

For the eigenvalue $\lambda_2 = 1 - ir = \overline{\alpha}$, we examine: $(B - (\overline{\alpha})I)\mathbf{v} = \mathbf{0}$, which is:

$$\begin{pmatrix} \overline{-\alpha} & -4 \\ 1 & \alpha \end{pmatrix} \mathbf{v} = \mathbf{0}, \text{ reducing to } \begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix} \mathbf{v} = \mathbf{0},$$

and the eigenspace is $Span\left(\left\{(\alpha, -1)^T\right\}\right)$.

We have $P^{-1}BP = diag(\alpha, \overline{\alpha})$,

$$\text{with } P = \begin{pmatrix} \overline{\alpha} & \alpha \\ -1 & -1 \end{pmatrix}, \quad \text{and so } P^{-1} = \frac{i}{2r} \begin{pmatrix} 1 & \alpha \\ -1 & -\overline{\alpha} \end{pmatrix}.$$

It now follows that,

$$B^k = PD^kP^{-1} = \begin{pmatrix} \overline{\alpha} & \alpha \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \overline{\alpha} \end{pmatrix}^k \frac{i}{2r} \begin{pmatrix} 1 & \alpha \\ -1 & -\overline{\alpha} \end{pmatrix}$$

$$= \begin{pmatrix} \overline{\alpha}\alpha^k & \alpha\overline{\alpha}^k \\ -\alpha^k & -\overline{\alpha}^k \end{pmatrix} \frac{i}{2r} \begin{pmatrix} 1 & \alpha \\ -1 & -\overline{\alpha} \end{pmatrix} = \frac{i}{2r} \begin{pmatrix} \overline{\alpha}\alpha^k - \overline{(\overline{\alpha}\alpha^k)} & \overline{(\alpha\overline{\alpha}^{k+1})} - \alpha\overline{\alpha}^{k+1} \\ \overline{\alpha}^k - \alpha^k & \overline{\alpha}^{k+1} - \alpha^{k+1} \end{pmatrix}.$$

If we recall for any complex number, z, we have $z - \overline{z} = 2i\, Im(z)$, then

$$B^k = \frac{1}{r} \begin{pmatrix} Im\overline{(\overline{\alpha}\alpha^k)} & Im(\alpha\overline{\alpha}^{k+1}) \\ Im(\alpha^k) & Im(\alpha^{k+1}) \end{pmatrix}.$$

We also have that: $\alpha = 1 + ir = 2\,cis(\frac{\pi}{3})$, $\overline{\alpha} = 2\,cis(\frac{-\pi}{3})$, $\alpha^k = 2^k\,cis(\frac{k\pi}{3})$, $\overline{\alpha}^k = 2^k\,cis(\frac{-k\pi}{3})$.

We then conclude that

$$B^k = \frac{2^k}{\sqrt{3}} \begin{pmatrix} 2\sin\left(\frac{(1-k)\pi}{3}\right) & 2^2\sin\left(\frac{-k\pi}{3}\right) \\ \sin\left(\frac{k\pi}{3}\right) & 2\sin\left(\frac{(1+k)\pi}{3}\right) \end{pmatrix}.$$

Topic 20

Special Subspaces and Bases

In this topic, we will revisit the concept of subspaces, and, in particular the important subspaces that we have encountered in this course, and we will consider finding bases for them.

We already know that if we have a subset U of $\mathbb{F}^n$, with $U = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, then $Span(U) = Span(\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\})$ is a subspace of $\mathbb{F}^n$.

Furthermore, many of the subspaces which we have encountered have naturally arisen as the span of some set of vectors, e.g. eigenspaces. It turns out that every subspace of $\mathbb{F}^n$ can be written this way, as we now show.

First of all we have to consider the trivial subspace of $\{\mathbf{0}\}$.

Definition 1: Basis and dimension for $\{\mathbf{0}\}$.

We say that $Span(\emptyset) = \{\mathbf{0}\}$, and that $\emptyset$ is a basis for $\{\mathbf{0}\}$.

Since there are zero vectors in $\emptyset$, we also say that the dimension of $\{\mathbf{0}\}$ is zero.

Lemma 1

Let V be a subspace of $\mathbb{F}^n$. Then there exists a linearly independent subset of V, W, with $p \leq n$ elements in it, such that

$$Span(W) = V.$$

Proof

We are dealing with $\mathbb{F}^n$, where n is a fixed positive integer.

We already know that a set of n linearly independent vectors forms a basis for $\mathbb{F}^n$, and that any subset of $\mathbb{F}^n$, with more than n vectors in it, is linearly dependent. Thus, any linearly independent subset of V is a linearly independent subset of $\mathbb{F}^n$, and thus has at most n vectors in it, that is $p \leq n$.

If $V = \{\mathbf{0}\}$, then $Span(\{\emptyset\}) = \{\mathbf{0}\}$, from Definition 1. Since $W = \emptyset \subset V$, and W has 0 vectors in it, the result is proven in this case.

If $V \neq \{\mathbf{0}\}$, then there exists $\mathbf{x} \in V$, such that $\mathbf{x} \neq \mathbf{0}$. Since V is a subspace of $\mathbb{F}^n$, it is closed under scalar multiplication and so $Span(\{\mathbf{x}\}) \subseteq V \subseteq \mathbb{F}^n$. There are two possibilities:

either, $Span(\{\mathbf{x}\}) = V$, then $\{\mathbf{x}\}$ is a linearly independent subset of V, with 1 vector in it and so the result is proven, $W = \{\mathbf{x}\}$,

or, $Span(\{\mathbf{x}\}) \neq V$, then there exists $\mathbf{y} \in V \backslash Span(\{\mathbf{x}\})$.

We now consider the set $Span(\{\mathbf{x}, \mathbf{y}\})$. This set is linearly independent as $\mathbf{y} \notin Span(\{\mathbf{x}\})$. We observe that since both $\mathbf{x}, \mathbf{y} \in V$, then $Span(\{\mathbf{x}, \mathbf{y}\}) \subseteq V$, since V is a subspace of $\mathbb{F}^n$, and is thus closed under addition and scalar multiplication. There are two possibilities:

either $V = Span(\{\mathbf{x}, \mathbf{y}\})$, and thus $W = \{\mathbf{x}, \mathbf{y}\}$, is a set which has 2 vectors in it. The result is then proven,

or $V \neq Span(\{\mathbf{x}, \mathbf{y}\})$, then there exists $\mathbf{z} \in V \backslash Span(\{\mathbf{x}, \mathbf{y}\})$.

And this process continues and must stop when W has at most n vectors in it. ■

Note that if W is a linearly independent subset of V which has n vectors in it, then W is a basis for $\mathbb{F}^n$, $Span(W) = \mathbb{F}^n$, and so $V = \mathbb{F}^n$.

Definition 2: Basis

Let U be a subspace of $\mathbb{F}^n$. We say that the subset W of U is a basis of U to mean that W is a linearly independent subset of U, with $Span(W) = U$.

Note that for W to be a basis for U we must have:

(a) $W \subseteq U$.

(b) W is linearly independent.

(c) $Span(W) = U$.

Example 1

Let $B = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$.

Find a basis for (i) the solution set to $B\mathbf{x} = \mathbf{0}$, and (ii) E_3.

Solution

In each case, we must first find the indicated subset (subspace), and then determine a basis for it.

(i) $B\mathbf{x} = \mathbf{0}$. We then have:

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \mathbf{x} = \mathbf{0},$$

which reduces to $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0},$

the solution set is $\left\{ s \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} : s \in \mathbb{F} \right\} = Span \left(\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \right).$

Notice that we automatically express the solution set as the span of W_1, with

$$W_1 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

Clearly, W_1 is linearly independent, thus it is a basis for the solution set to $B\mathbf{x} = \mathbf{0}$.

(ii) E_3. We then have:

$(B - 3I)\,\mathbf{x} = \mathbf{0}$, which is $\begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \mathbf{x} = \mathbf{0},$

which reduces to $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0},$

and the solution set is $\left\{ a \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} : a, b \in \mathbb{F} \right\}$

$$= Span \left(\left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\} \right).$$

Notice that we automatically express the solution set as the span of W_2, with

$$W_2 = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}.$$

Clearly, W_2 is linearly independent, thus it is a basis for the eigenspace E_3.

Lemma 2: Number of vectors in a basis of a subspace V of $\mathbb{F}^n$

Let V be a subspace of $\mathbb{F}^n$, and let $U = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be a basis for V.

$$\text{If } W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_q\} \text{ be a basis for } V, \text{ then } p = q.$$

Proof

Let $W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_q\}$ be a basis for V.

We already know from Lemma 1 that $p \leq n$ and $q \leq n$.

Let us form the two matrices:

$$A = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p) \in M_{n\times p}(\mathbb{F}) \;\text{ and }\; B = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_q) \in M_{n\times q}(\mathbb{F}).$$

Since the two sets U and W are linearly independent, we know that:

$$rank(A) = p \;\text{ and }\; rank(B) = q.$$

Since U is a basis for V and $W \subset V$, then each vector is W can be expressed as a linear combination of the vectors in U, that is, there exist scalars c_{ji} (not c_{ij}) with

$$\mathbf{w}_i = \sum_{j=1}^{j=p} c_{ji}\,\mathbf{u}_j, \quad \text{for } i = 1, \ldots, q.$$

If we construct a matrix $C \in M_{p\times q}(\mathbb{F})$ defined by $(C)_{ji} = c_{ji}$, then we get that $B = AC$.

We now show that the columns of C, which are columns in $\mathbb{F}^p$, are linearly independent.

If $C\,\mathbf{x} = \mathbf{0}$, then $AC\,\mathbf{x} = \mathbf{0}$, and so $B\,\mathbf{x} = \mathbf{0}$. However the columns of B are linearly independent and so the only solution of this equation is $\mathbf{x} = \mathbf{0}$, and we conclude that the columns of C are linearly independent. Thus $q \leq p$.

We can repeat this discussion, starting, instead, by writing each vector in U as a linear combination of the vectors in W. The result would be that $p \leq q$.

We then conclude that $p = q$. ■

Definition 3: Dimension

Let V be a subspace of $\mathbb{F}^n$, we say that the **dimension** of V is p, $dim(V) = p$ to mean that there are p vectors in a basis for V.

Example 2

Returning to Example 1, where we considered the matrix $B = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$.

The solution set to $B\,\mathbf{x} = \mathbf{0}$ was found to be $Span\,(W_1)$, with $W_1 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$.

The dimension of the solution set to $B\mathbf{x} = \mathbf{0}$, is 1.

In addition, $E_3 = Span\,(W_2)$, where $W_2 = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$.

We conclude that $\dim(E_3) = 2$.

Suppose we have a basis W, for a subspace V of dimension k, of $\mathbb{F}^n$. We wish to continue to use the vectors in this basis W, but we now need a basis, B, for $\mathbb{F}^n$.
How do we extend W to obtain B? One way to proceed is to use the Replacement Theorem, in which we essentially replace k of the vectors in the standard basis S of $\mathbb{F}^n$, with the vectors in W.

Lemma 3: (Replacement Theorem)

Let V be a subspace of $\mathbb{F}^n$ such that $dim(V) = k > 0$.

If W is a basis for V, then it is possible to extend W to a basis B of $\mathbb{F}^n$.

Proof

We will prove this result in a constructive manner.
Let $W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ and let S is the standard basis for $\mathbb{F}^n$.
Let us examine the set $U = W \cup S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k+n}\}$, with $\mathbf{u}_i = \mathbf{w}_i$, $i = 1, \ldots, k$.

Since S is a basis for $\mathbb{F}^n$, then $Span(S) = Span(U) = \mathbb{F}^n$. We thus have a spanning set for $\mathbb{F}^n$: however, it is linearly dependent as it contains more than n vectors.

We now apply Lemma 5 in Topic 17B, in which we reduce a spanning set to a linearly independent set with the same span. Furthermore, since W is a basis for V, and is thus linearly independent, then the first k columns of the matrix involved in this reduction process must be pivot columns. The remaining pivot columns will determine which vectors to select from S. The k columns that are not pivot columns correspond to the vectors from S that are replaced by the k vectors from W, hence the name “Replacement Theorem”. ■

Example 3

Extend the basis for the eigenspace E_3 from Example 2, into a basis for $\mathbb{R}^3$.

Solution

We already have $E_3 = Span\,(W)$, where $W = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$ is a basis for E_3.

Let $U = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$.

We now follow the technique of Lemma 5 in Topic 17B, and reduce this set to a linearly independent set.

Consider the linear combination:

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5 = \mathbf{0}.$$

The coefficient matrix is $\begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix}$ which reduces to $\begin{pmatrix} 1 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$.

We conclude that the first three vectors are linearly independent.

Thus a basis for $\mathbb{R}^3$, which has the basis W_2 for E_3 as a subset is:

$$B = \left\{ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

There are many other choices of course.

Example 4

Suppose that $W = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ is a basis for a subspace of $\mathbb{R}^5$.

Extend this basis into a basis for $\mathbb{R}^5$.

Solution

We consider the set

$$U = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5, \mathbf{u}_6, \mathbf{u}_7, \mathbf{u}_8\}$$

$$= \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

We now follow the technique of Lemma 5 in Topic 17B, and reduce this set to a linearly independent set. Consider the linear combination:

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5 + c_6\mathbf{u}_6 + c_7\mathbf{u}_7 + c_8\mathbf{u}_8 = \mathbf{0}.$$

The coefficient matrix is

$$\begin{pmatrix} 1 & 5 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 & 0 & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 & 0 & 0 \\ 4 & 2 & 1 & 0 & 0 & 0 & 1 & 0 \\ 5 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ which reduces to } \begin{pmatrix} 1 & 0 & 0 & \frac{1}{6} & 0 & 0 & -\frac{5}{6} & \frac{5}{6} \\ 0 & 1 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & -\frac{1}{6} \\ 0 & 0 & 1 & -1 & 0 & 0 & 4 & -3 \\ 0 & 0 & 0 & 0 & 1 & 0 & -3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

We conclude that the vectors $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_5, \mathbf{u}_6\}$ are linearly independent.

Thus a basis for $\mathbb{R}^5$ which has the basis W of a subspace of $\mathbb{R}^5$ as a subset, is:

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

There are many other choices of course.

Example 5

Find a basis for the column space of the matrix $\begin{pmatrix} 1 & 5 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 & 0 & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 & 0 & 0 \\ 4 & 2 & 1 & 0 & 0 & 0 & 1 & 0 \\ 5 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$, in Ex. 4.

Solution

We have reduced this matrix in Example 4, and we know that the pivot columns are columns 1, 2, 3, 5 and 6.

Thus the corresponding columns form a basis for the column space of this matrix.

Example 6

Find a basis for the column space of the matrix $A = \begin{pmatrix} 1 & 2 & 1 & 3 & 3 & -2 \\ 2 & 4 & 1 & 4 & 2 & -4 \\ 3 & 6 & 1 & 5 & 1 & -7 \end{pmatrix}$.

Solution

We row reduce A to get : $\begin{pmatrix} 1 & 2 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$.

The pivot columns are columns 1, 3 and 6. And thus we conclude that a basis for the column space is:

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

Remark 1: $rank(A) = \dim(\mathrm{Col}(A))$.

Let $A \in M_{m\times n}(\mathbb{F})$. Then the rank of A is equal to the dimension of the column space of A. This follows since the *rank* of A is equal to the number of pivots that A has, and each pivot column provides a linearly independent column to the column space of A.

Remark 2: There are two subspaces for which we often obtain a basis naturally when we find them, and these are the solution set to a system of homogeneous linear equations, and eigenspaces.

From an arithmetical point of view, these are identical, as, if $A \in M_{n\times n}(\mathbb{F})$, with eigenvalue λ, then $E_\lambda = \{(A - \lambda I)\mathbf{x} = \mathbf{0}\}$. That is, to find an eigenspace, you must solve a homogeneous linear system.

Remark 3: Recall that the solution set to a system of inhomogeneous linear equations is not a subspace.

Example 7

We revisit an Example 4 in Topic 10A:

$$\begin{cases} x - 2y - z + 3w = 0 \\ \qquad 2x - 4y + z = 0 \\ x - 2y + 2z - 3w = 0 \end{cases} .$$

The reduced row echelon form of the coefficient matrix is:

$$\left(\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Letting $y = p \in \mathbb{R}$ and $w = q \in \mathbb{R}$, then $x = 3 + 2p - q,$ and $z = 2 + 2q$.

We can thus write the solution set U as:

$$U = \left\{ \begin{pmatrix} 2p - q \\ p \\ 2q \\ q \end{pmatrix} : \ p, q \in \mathbb{R} \right\} = \left\{ p \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + q \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} : \ p, q \in \mathbb{R} \right\}$$

$$= Span \left(\left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \right\} \right) = Span(W).$$

Note that we automatically wrote the solution set as the span of a set W of some vectors.

In addition, it is clear that the set W is linearly independent. This is due to the position of the zeros in the vectors in the set W. This is a feature of every homogeneous linear system that we have solved in this course: each different parameter in the solution set corresponds to a vector in the solution set, and it will be independent of the vectors provided by any and all the other parameters, as this vector will have a non-zero entry (usually a 1) in a row in which all the other vectors have a zero.

We conclude by restating a Lemma 2 in Topic 9, which allowed us to use the rank of a matrix to determine the number of parameters in the solution set to a homogeneous linear system of equations. The only new feature is that we can now refer to the dimensions of the various subspaces that occur.

Theorem 1: The Dimension Theorem (or Rank-Nullity Theorem).

Let $A \in M_{m\times n}(\mathbb{F})$. Then

$$n = \dim(\mathrm{Col}(A)) + \dim(N(A)).$$

This can be written as:

$$n = rank(A) + nullity(A),$$

$$n = rank(T_A) + nullity(T_A).$$

Topic 21A

Vector Space I

In this course, all of our investigations have taken place within $\mathbb{R}^n$ and $\mathbb{C}^n$ or within various subspaces of these spaces. When considering any one of these spaces, referred to generically as V, we have benefited from our ability to be able to combine our vectors in V through the operations of addition and scalar multiplication. We have taken it for granted that the construction of such linear combinations of vectors in V will always yield another vector in V, that is, closure under additions and scalar multiplication is a fundamental feature of the linear algebra world.

To be quite explicit, when we are dealing with V,

(a) we have an addition operation which allows us to combine two vectors in V, and produce another vector in V,

(b) we also have a scalar multiplication law, which allows us to use another set of objects $\mathbb{F}$, called scalars, and we can scale a vector in V by a scalar in $\mathbb{F}$, to produce another vector in V.

There are many other area in mathematics in which we have a similar underlying structure, for example, suppose we consider the space of real quadratic polynomials,

$$P_2(\mathbb{R}) = \left\{a + bx + cx^2 : a, b, c \in \mathbb{R}\right\}.$$

Notice that if you take two quadratic polynomials,

$$p_1(x) = a_1 + b_1x + c_1x^2 \ \text{ and } \ p_2(x) = a_2 + b_2x + c_2x^2,$$

and add them up, equating coefficients, then we have:

$$\begin{aligned} p_1(x) + p_2(x) = p_3(x) &= a_1 + b_1x + c_1x^2 + a_2 + b_2x + c_2x^2 \\ &= a_3 + b_3x + c_3x^2, \quad \text{with } a_3 = a_1 + a_2,\ b_3 = b_1 + b_2,\ c_3 = c_1 + c_2, \end{aligned}$$

which is another (at most) quadratic polynomial.

If you scale a quadratic polynomial by a real number d, then you obtain:

$$\begin{aligned} dp_1(x) = p_4(x) &= d(a_1 + b_1x + c_1x^2) = da_1 + db_1x + dc_1x^2 \\ &= a_4 + b_4x + c_4x^2, \quad \text{with } a_4 = da_1,\ b_4 = db_1,\ c_4 = dc_1, \end{aligned}$$

which is another (at most) quadratic polynomial.

Thus $P_2(\mathbb{R})$ is closed under addition and scalar multiplication.

However, there are a number of other features lurking in the background of the spaces which we have been dealing with in our linear algebra course. We now expose them all.

Vector Space - Definition

We can think of linear algebra as a game in which I must provide you with four objects: a set V, an addition law, which we denote by $\oplus$ (since it may not be the one that you are used to), another set of scalars ($\mathbb{F}$, in this course), and a scalar multiplication law $\odot$.

(I) Addition is our way of combining the elements of V: we denote it by $\oplus$, and when we perform addition of two elements of V, we want to obtain an element of the set V.

$$\text{For any } \mathbf{v}, \mathbf{w} \in V,\ \mathbf{v} \oplus \mathbf{w} \text{ is defined and } \mathbf{v} \oplus \mathbf{w} \in V.$$

(II) Scalar multiplication is our way of combining an element in V by an element in $\mathbb{F}$: we denote it by $\odot$, and when we scale an element of V by a scalar, we want to obtain an element of the set V.

$$\text{For any } \mathbf{v} \in V,\ \text{and for any } c \in \mathbb{F},\ c \odot \mathbf{v} \text{ is defined and } c \odot \mathbf{v} \in V.$$

We say then that the set V is closed under addition and scalar multiplication.

There are another eight axioms that must also be satisfied for the set V, with its addition law and scalar multiplication law by elements of $\mathbb{F}$, in order to have a vector space.

These are:

(a) For all $\mathbf{v}, \mathbf{w} \in V,\ \mathbf{v} \oplus \mathbf{w} = \mathbf{w} \oplus \mathbf{v}$.

(b) For all $\mathbf{v}, \mathbf{w}, \mathbf{z} \in V,\ (\mathbf{v} \oplus \mathbf{w}) \oplus \mathbf{z} = \mathbf{v} \oplus (\mathbf{w} \oplus \mathbf{z}) = \mathbf{v} \oplus \mathbf{w} \oplus \mathbf{z}$.

(c) There exists a vector $\mathbf{0} \in \mathbf{V}$ for which $\mathbf{v} \oplus \mathbf{0} = \mathbf{0} \oplus \mathbf{v} = \mathbf{v}$, for all $\mathbf{v} \in V$.

(d) For all $\mathbf{v} \in V,$ there exists a vector, $(-\mathbf{v}) \in V$, such that $\mathbf{v} \oplus (-\mathbf{v}) = (-\mathbf{v}) \oplus \mathbf{v} = \mathbf{0}$.

(e) For all $\mathbf{v}, \mathbf{w} \in V$, and for all $c \in \mathbb{F},\ c \odot (\mathbf{v} \oplus \mathbf{w}) = (c \odot \mathbf{v}) \oplus (c \odot \mathbf{w})$.

(f) For all $\mathbf{v} \in V$, and for all $c, d \in \mathbb{F},\ (c + d) \odot \mathbf{v} = (c \odot \mathbf{v}) \oplus (d \odot \mathbf{v})$, where $+$ is the addition in $\mathbb{F}$.

(g) For all $\mathbf{v} \in V$, for all $c, d \in \mathbb{F},\ (c \times d) \odot \mathbf{v} = c \odot (d \odot \mathbf{v})$, where $\times$ is multiplication in $\mathbb{F}$.

(h) For all $\mathbf{v} \in V$, $1 \odot \mathbf{v} = \mathbf{v}$.

Definition 1: Vector Space

Suppose we are given a set V, an addition law $\oplus$, a field $\mathbb{F}$, and a scalar multiplication law $\odot$, for which the two closure axioms (I) and (II), and the eight axioms (a),...,(h) hold.

Then we say that $(V, \oplus, \mathbb{F}, \odot)$ is a **vector space** over $\mathbb{F}$, with the given addition law $\oplus$, and the given scalar multiplication law $\odot$. And the elements of V are called vectors.

Our prototype vector spaces are $\mathbb{R}^n$ and $\mathbb{C}^n$, but we have met a number of other ones during the progress of this course.

Example 1: Vector Spaces

(A) Consider the set $V = M_{m\times n}(\mathbb{F})$ with addition of matrices, $\oplus$, defined in the usual way and scalar multiplication by scalars from $\mathbb{F}$, $\odot$, defined in the standard way. Then V, is a vector space over $\mathbb{F}$.
The zero vector in this space is the zero matrix.

(B) Consider the set V of all linear transformations from $\mathbb{F}^n$ to $\mathbb{F}^m$, with addition, $\oplus$, defined in the usual way for functions and scalar multiplication by scalars from $\mathbb{F}$, $\odot$, defined in the standard way of multiplying a function by a scalar. Then V, is a vector space over $\mathbb{F}$.
The zero vector is the linear transformation $T_0 : \mathbb{F}^n \to \mathbb{F}^m$, with $T_0(\mathbf{x}) = \mathbf{0}_m$.

(C) Consider the set $V = P_n(\mathbb{F})$, of all polynomials with coefficients in $\mathbb{F}$, and of order at most n. We define addition, $\oplus$, in the standard way, see the example in $P_2(\mathbb{R})$ above, and scalar multiplication by scalars from $\mathbb{F}$, $\odot$, defined in the standard way of multiplying a function by a scalar. Then V, is a vector space over $\mathbb{F}$.
The zero vector is the zero polynomial in $P_n(x)$, namely $p_0(x) = 0$ or more simply the function $y = 0$.

Example 2: A Weird Vector Space

Consider the set $\mathbb{R}^2$ with addition $\oplus$ defined by

$$\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a+c \\ b+d+1 \end{pmatrix},$$

and scalar multiplication $\odot$ defined by

$$k \odot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} k \times a \\ (k \times b) + k - 1 \end{pmatrix} = \begin{pmatrix} ka \\ kb + k - 1 \end{pmatrix}.$$

Prove that $(\mathbb{R}^2, \oplus, \mathbb{R}, \odot)$ is a vector space.

Solution

We have closure of addition and of scalar multiplication as the objects $(a+c)$, $(b+d+1)$, (ka) and $(kb+k-1)$ are in $\mathbb{R}$.

We check the eight properties.

(a) $\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a+c \\ b+d+1 \end{pmatrix}$

$\begin{pmatrix} c \\ d \end{pmatrix} \oplus \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} c+a \\ d+b+1 \end{pmatrix} = \begin{pmatrix} a+c \\ b+d+1 \end{pmatrix}$, since addition in $\mathbb{R}$ is commutative.

We conclude that

$$\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix} \oplus \begin{pmatrix} a \\ b \end{pmatrix}, \quad \text{for all } \begin{pmatrix} a \\ b \end{pmatrix} \text{ and } \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{R}^2$$

.

(b) $\left(\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c \\ d \end{pmatrix}\right) \oplus \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} a+c \\ b+d+1 \end{pmatrix} \oplus \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} (a+c)+e \\ (b+d+1)+f+1 \end{pmatrix}$

$\begin{pmatrix} a \\ b \end{pmatrix} \oplus \left(\begin{pmatrix} c \\ d \end{pmatrix} \oplus \begin{pmatrix} e \\ f \end{pmatrix}\right) = \begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c+e \\ d+f+1 \end{pmatrix} = \begin{pmatrix} a+(c+e) \\ b+(d+f+1)+1 \end{pmatrix}$

and $\begin{pmatrix} (a+c)+e \\ (b+d+1)+f+1 \end{pmatrix} = \begin{pmatrix} a+(c+e) \\ b+(d+f+1)+1 \end{pmatrix} = \begin{pmatrix} a+c+e \\ b+d+f+2 \end{pmatrix}$,

as addition in $\mathbb{R}$ is associative.

We conclude that for all $\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix}, \begin{pmatrix} e \\ f \end{pmatrix} \in \mathbb{R}^2$,

$$\left(\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c \\ d \end{pmatrix}\right) \oplus \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \oplus \left(\begin{pmatrix} c \\ d \end{pmatrix} \oplus \begin{pmatrix} e \\ f \end{pmatrix}\right).$$

(c) Since $\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} a+0 \\ b-1+1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$, for all $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$,

we conclude that the zero vector of $(\mathbb{R}^2, \oplus, \mathbb{R}, \odot)$ is $\mathbf{0} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

(d) Since $\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} -a \\ -b-2 \end{pmatrix} = \begin{pmatrix} a-a \\ b-b-2+1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, for all $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$, we conclude that the additive inverse in $(\mathbb{R}^2, \oplus, \mathbb{R}, \odot)$ of $\begin{pmatrix} a \\ b \end{pmatrix}$ is $\begin{pmatrix} -a \\ -b-2 \end{pmatrix}$, that is,

$$-\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -a \\ -b-2 \end{pmatrix}.$$

(e) $k \odot \left(\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c \\ d \end{pmatrix} \right) = k \odot \begin{pmatrix} a+c \\ b+d+1 \end{pmatrix} = \begin{pmatrix} k(a+c) \\ k(b+d+1)+k-1 \end{pmatrix}$

$$= \begin{pmatrix} k(a+c) \\ k(b+d)+2k-1 \end{pmatrix}$$

$$\left(k \odot \begin{pmatrix} a \\ b \end{pmatrix}\right) \oplus \left(k \odot \begin{pmatrix} c \\ d \end{pmatrix}\right) = \begin{pmatrix} ka \\ kb+k-1 \end{pmatrix} \oplus \left(\begin{pmatrix} kc \\ kd+k-1 \end{pmatrix}\right)$$

$$= \begin{pmatrix} ka+kc \\ kb+k-1+kd+k-1+1 \end{pmatrix} = \begin{pmatrix} k(a+c) \\ k(b+d)+2k-1 \end{pmatrix},$$

since addition and multiplication in $\mathbb{R}$ are distributive.

We conclude that for all $\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{R}^2$, and for all $k \in \mathbb{R}$,

$$k \odot \left(\begin{pmatrix} a \\ b \end{pmatrix} \oplus \begin{pmatrix} c \\ d \end{pmatrix} \right) = \left(k \odot \begin{pmatrix} a \\ b \end{pmatrix}\right) \oplus \left(k \odot \begin{pmatrix} c \\ d \end{pmatrix}\right).$$

(f) $(k+l) \oplus \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} (k+l)a \\ (k+l)b+k+l-1 \end{pmatrix}$

$$\left(k \odot \begin{pmatrix} a \\ b \end{pmatrix}\right) \oplus \left(l \odot \begin{pmatrix} a \\ b \end{pmatrix}\right) = \begin{pmatrix} ka \\ kb+k-1 \end{pmatrix} \oplus \begin{pmatrix} la \\ lb+l-1 \end{pmatrix},$$

$$= \begin{pmatrix} (k+l)a \\ (k+l)b+k-1+l-1+1 \end{pmatrix} = \begin{pmatrix} (k+l)a \\ (k+l)b+k+l-1 \end{pmatrix}.$$

We conclude that for all $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$ and for all $k, l \in \mathbb{R}$, that

$$(k+l) \oplus \begin{pmatrix} a \\ b \end{pmatrix} = \left(k \odot \begin{pmatrix} a \\ b \end{pmatrix}\right) \oplus \left(l \odot \begin{pmatrix} a \\ b \end{pmatrix}\right).$$

(g) $(k \times l) \odot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} kla \\ klb+kl-1 \end{pmatrix}$

$$k \odot \left(l \odot \begin{pmatrix} a \\ b \end{pmatrix} \right) = k \odot \begin{pmatrix} la \\ lb + l - 1 \end{pmatrix} = \begin{pmatrix} kla \\ k(lb + l - 1) + k - 1 \end{pmatrix}$$

$$= \begin{pmatrix} kla \\ klb + kl - 1 \end{pmatrix}.$$

We conclude that for all $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$ and for all $k, l \in \mathbb{R}$, that

$$(k \times l) \odot \begin{pmatrix} a \\ b \end{pmatrix} = k \odot \left(l \odot \begin{pmatrix} a \\ b \end{pmatrix} \right).$$

(h) Since $1 \odot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \times a \\ (1 \times b) + 1 - 1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$,

We conclude that for all $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$ that $1 \odot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$.

And thus we have proved that $(\mathbb{R}^2, \oplus, \mathbb{R}, \odot)$ is a vector space.

This example is just to warn you that you cannot necessarily take things for granted in a vector space, here the zero vector and negative inverse are rather unusual.
We will not spend time dealing with such esoteric vector spaces, but they do exist.

Once we have a vector space, we are interested in performing all the familiar operations and investigations with them.

In these few lectures will will only have to time to introduce some of the more important ideas.

Definition 2: Linear Combination

Let $(V, \oplus, \mathbb{F}, \odot)$ be a vector space over $\mathbb{F}$, with $\mathbf{v}_1$ and $\mathbf{v}_2$ in V, and c_1 and c_2 in $\mathbb{F}$.

Then we refer to

$$(c_1 \odot \mathbf{v}_1) \oplus (c_2 \odot \mathbf{v}_2)$$

as a **linear combination** of $\mathbf{v}_1$ and $\mathbf{v}_2$, or as a linear combination of elements of V.

Note that under closure of the vector space, a linear combination of elements of V will be another vector in V.

Note, also, that we can extend the concept of linear combination to any finite number of vectors that lie in V.

Example 3: Linear combinations

Consider the vector space $P_2(\mathbb{R})$ over $\mathbb{R}$, with the usual addition of polynomials and the usual scalar multiplication.

Then $p(x) = 2(1-2x+3x^2) - 3(9-5x+x^2)$ is a linear combination of vectors of $P_2(\mathbb{R})$.

Consider the vector space $M_{3\times 2}(\mathbb{R})$ over $\mathbb{R}$, with the usual addition of matrices and the usual scalar multiplication.

Then $A = 4\begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix} - 2\begin{pmatrix} -5 & 9 \\ -7 & 7 \\ -9 & 5 \end{pmatrix} + 3\begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$, is a linear combination of vectors in $M_{3\times 2}(\mathbb{R})$.

Definition 3: Span

Let $(V, \oplus, \mathbb{F}, \odot)$ be a vector space over $\mathbb{F}$, and let $W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\} \subset V$.

We define the Span of W, denoted by $Span(W)$, to be the set of all linear combinations of elements of W, that is:

$$Span(W) = \{(c_1 \odot \mathbf{w}_1) \oplus (c_2 \odot \mathbf{w}_2) \oplus \cdots \oplus (c_p \odot \mathbf{w}_p) : c_i \in \mathbb{F},\ i = 1, \ldots, p\}.$$

Example 4: Span

Consider the vector space $P_2(\mathbb{R})$ over $\mathbb{R}$, with the usual addition of polynomials and the usual scalar multiplication.

Then

$$Span(\{(1-2x+3x^2), (9-5x+x^2)\})$$

$$= \left\{c_1(1-2x+3x^2) + c_2(9-5x+x^2) : c_1, c_2 \in \mathbb{R}\right\}.$$

Consider the vector space $M_{3\times 2}(\mathbb{R})$ over $\mathbb{R}$, with the usual addition of matrices and the usual scalar multiplication.

$$Span\left(\left\{\begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix}, \begin{pmatrix} -5 & 9 \\ -7 & 7 \\ -9 & 5 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}\right\}\right)$$

$$= \left\{ c_1 \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix} + c_2 \begin{pmatrix} -5 & 9 \\ -7 & 7 \\ -9 & 5 \end{pmatrix} + c_3 \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix} : c_1, c_2, c_3 \in \mathbb{R} \right\}.$$

Equations

We can consider the concept of equality in a vector space, and just as in $\mathbb{F}^n$, this will lead to a system of linear equations, which we can solve with the usual matrix related techniques.

We use the technique of comparing coefficients, which really relies on the concepts of a basis and linear independence.

We will talk about these ideas in the next lecture.

Example 5

Consider the vector space $P_2(\mathbb{R})$ over $\mathbb{R}$, with the usual addition of polynomials and the usual scalar multiplication.

Is the vector $p(x) = 1 + 2x + 3x^2 \in Span(\{(1 - 2x + 3x^2), (9 - 5x + x^2)\})$?

Solution

The questions is thus whether we can find constants $a, b \in \mathbb{R}$, such that

$$1 + 2x + 3x^2 = a(1 - 2x + 3x^2) + b(9 - 5x + x^2) \;\Leftrightarrow$$

$$1 + 2x + 3x^2 = (1a + 9b) + (-2a - 5b)x + (3a + b)x^2.$$

Comparing coefficients yields the following system of three equations:

$$\left\{ \begin{aligned} a + 9b &= 1 \\ -2a - 5b &= 2 \\ 3a + b &= 3 \end{aligned} \right. .$$

The augmented matrix for this system is:

$$\left(\begin{array}{rr|r} 1 & 9 & 1 \\ -2 & -5 & 2 \\ 3 & 1 & 3 \end{array} \right), \quad \text{which row reduces to} \quad \left(\begin{array}{rr|r} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right).$$

Since the last column of the augmented matrix is a pivot column, then the system is inconsistent and thus

$$p(x) = 1 + 2x + 3x^2 \notin Span(\{(1 - 2x + 3x^2), (9 - 5x + x^2)\}).$$

Consider the vector space $M_{3\times 2}(\mathbb{R})$ over $\mathbb{R}$, with the usual addition of matrices and the usual scalar multiplication.

Is $\begin{pmatrix} 17 & -13 \\ 16 & -4 \\ 31 & -11 \end{pmatrix} \in Span\left(\left\{\begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix}, \begin{pmatrix} -5 & 9 \\ -7 & 7 \\ -9 & 5 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}\right\}\right)$?

Solution

The question is whether we can find real constants, a, b, c such that

$$\begin{pmatrix} 17 & -13 \\ 16 & -4 \\ 31 & -11 \end{pmatrix} = a\begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix} + b\begin{pmatrix} -5 & 9 \\ -7 & 7 \\ -9 & 5 \end{pmatrix} + c\begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix} \Leftrightarrow$$

$$\begin{pmatrix} 17 & -13 \\ 16 & -4 \\ 31 & -11 \end{pmatrix} = \begin{pmatrix} a - 5b + c & 3a + 9b - c \\ 2a - 7b - c & 2a + 7b + c \\ 3a - 9b + c & a + 5b - c \end{pmatrix}.$$

Comparing coefficients yields the following system of six equations:

$$\begin{cases} a - 5b + c = 17 \\ 2a - 7b - c = 16 \\ 3a - 9b + c = 31 \\ 3a + 9b - c = -13 \\ 2a + 7b + c = -4 \\ a + 5b - c = -11 \end{cases}.$$

The augmented matrix for this system is

$$\begin{pmatrix} 1 & -5 & 1 & | \, 17 \\ 2 & -7 & -1 & | \, 16 \\ 3 & -9 & 1 & | \, 31 \\ 3 & 9 & -1 & | -13 \\ 2 & 7 & 1 & | -4 \\ 1 & 5 & -1 & | -11 \end{pmatrix}, \text{ which row reduces to } \begin{pmatrix} 1 & 0 & 0 & | \, 3 \\ 0 & 1 & 0 & | -2 \\ 0 & 0 & 1 & | \, 4 \\ 0 & 0 & 0 & | \, 0 \\ 0 & 0 & 0 & | \, 0 \\ 0 & 0 & 0 & | \, 0 \end{pmatrix}.$$

The system is consistent. We conclude that the answer is yes, with the constants in the linear combination being: $a = 3$, $b = -2$, $c = 4$.

And thus $\begin{pmatrix} 17 & -13 \\ 16 & -4 \\ 31 & -11 \end{pmatrix} \in Span\left(\left\{\begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix}, \begin{pmatrix} -5 & 9 \\ -7 & 7 \\ -9 & 5 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}\right\}\right)$.

Definition 4: Vector subspace

Let $(V, \oplus, \mathbb{F}, \odot)$ be a vector space and let U be a subset of V.

We say that U is a vector subspace of V to mean that U is a non-empty subset that is closed under addition and scalar multiplication.

Thus, if the subset U of V is a vector subspace of V, then

(i) $U \neq \emptyset$,

(ii) $\forall \mathbf{u}_1, \mathbf{u}_2 \in U,\ \mathbf{u}_1 \oplus \mathbf{u}_2 \in U$,

(iii) $\forall \mathbf{u}_1 \in U$ and $\forall c \in \mathbb{F}$, $c \odot \mathbf{u}_1 \in U$.

In practice, when we examine the subset, we check if the zero vector is in it, if it is not, then the subset cannot be a subspace.

Example 6

Consider the vector space $P_2(\mathbb{R})$ over $\mathbb{R}$, with the usual addition of polynomials and the usual scalar multiplication.

Which of the following subsets of $P_2(\mathbb{R})$ are vector subspaces of $P_2(\mathbb{R})$?

$S_1 = \{p(x) \in P_2(\mathbb{R}) : p(0) + p(1) = 2\}$.

$S_2 = \{p(x) \in P_2(\mathbb{R}) : p(0)p(1) = 0\}$.

$S_3 = \{p(x) \in P_2(\mathbb{R}) : p(0) + p'(1) = 0\}$.

Solution

S_1 is not a subspace since the zero vector $(p_0(x) = 0)$, is not in S_1, since $p_0(0) + p_0(1) = 0 \neq 2$.

The zero vector is in S_2, since $p_0(0)p_0(1) = 0$, however, we might suspect that this set is not closed under addition.

Let $p_1(x) = x$ and $p_2(x) = 1 - x$, then both $p_1(x)$ and $p_2(x)$ are in S_2.

However, $p_1(x)+ p_2(x) = p_3(x) = 1$, but $p_3(x) \notin S_2$.

We conclude that S_2 is not a subspace.

$S_3 = \{p(x) \in P_2(\mathbb{R}) : p(0) + p'(1) = 0\}$. This is a subspace of $P_2(\mathbb{R})$.

(i) the zero vector is in S_3, since if $p_0(x) = 0$, then $p_0'(x) = 0$ and $p_0(0) + p_0'(1) = 0 + 0 = 0$.

Let $p_1(x) = a + bx + cx^2 \in S_3$, then $a + b + 2c = 0$.

Let $p_2(x) = d + ex + fx^2 \in S_3$, then $d + e + 2f = 0$.

Consider $p_3(x) = p_1(x) + p_2(x) = a + bx + cx^2 + d + ex + fx^2 = (a + d) + (b + e)x + (c + f)x^2$.

Then $p_3(0)+p_3'(1) = (a+d)+(b+e)+2(c+f) = a+b+2c+d+e+2f = 0$. So that $p_3(x) \in S_3$.

Also, $p_4(x) = k\, p_1(x) = k(a + bx + cx^2)$, then

$p_4(0) + p_4'(1) = ka + k(b + 2c) = k(a + b + 2c) = 0$. So that $p_4(x) \in S_3$.

We conclude that S_3 is a subspace of $P_2(\mathbb{R})$.

Topic 21B

Vector Space - II

In this lecture, we will be examine some of the features of generic vector spaces. We commence with a number of rather unsurprising lemmas.

Throughout this lecture, we assume that $(V, \oplus, \mathbb{F}, \odot)$ is a vector space.

Lemma 1

The zero vector in a vector space V is unique.

Proof

Let us suppose that there exist $\mathbf{0} \in V$, and $\tilde{\mathbf{0}} \in \mathbf{V}$, both with the property that

(i) $\mathbf{v} \oplus \mathbf{0} = \mathbf{v} = \mathbf{0} \oplus \mathbf{v}$, $\forall \mathbf{v} \in V$, and

(ii) $\mathbf{v} \oplus \tilde{\mathbf{0}} = \mathbf{v} = \tilde{\mathbf{0}} \oplus \mathbf{v}$, $\forall \mathbf{v} \in V$.

Then we have that $\tilde{\mathbf{0}} = \mathbf{0} \oplus \tilde{\mathbf{0}}$, using (i), and so $\tilde{\mathbf{0}} = \mathbf{0} \oplus \tilde{\mathbf{0}} = \mathbf{0}$, using (ii). ■

Lemma 2

The additive inverse $(-\mathbf{x})$ of a vector $\mathbf{x}$ in a vector space V, is unique.

Proof

Let us suppose that there exist $\mathbf{y} \in V$, and $\mathbf{z} \in \mathbf{V}$, both with the property that

(i) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} = \mathbf{0}$, and

(ii) $\mathbf{x} + \mathbf{z} = \mathbf{z} + \mathbf{x} = \mathbf{0}$.

Then we have that $\mathbf{y} + \mathbf{x} + \mathbf{z} = (\mathbf{y} + \mathbf{x}) + \mathbf{z} = \mathbf{y} + (\mathbf{x} + \mathbf{z})$, using axiom (b) and so $\mathbf{0} + \mathbf{z} = \mathbf{y} + \mathbf{0}$, that is $\mathbf{z} = \mathbf{y}$. ■

Lemma 3: Properties of zero

Let V be a vector space and $x \in V$. We then have that

$$0 \odot \mathbf{x} = \mathbf{0}, \;\; \forall \mathbf{x} \in V, \quad \text{and} \;\; a \odot \mathbf{0} = \mathbf{0}, \;\; \forall a \in \mathbb{F}.$$

Proof: see Topic 1.

We denote the additive inverse of the vector $\mathbf{x}$ by $(-\mathbf{x})$.

Lemma 4: The additive inverse

Let $(-\mathbf{x})$ be the additive inverse of a vector $\mathbf{x}$ in a vector space V. Then,

$$(-\mathbf{x}) = (-1) \odot \mathbf{x}.$$

Proof

Using axiom (f), we get that $((-1) \odot \mathbf{x}) \oplus ((1) \odot \mathbf{x}) = ((-1) + 1) \odot \mathbf{x} = 0 \odot \mathbf{x} = \mathbf{0}$.

Since $((-1) \odot \mathbf{x}) \oplus ((1) \odot \mathbf{x}) = ((-1) \odot \mathbf{x}) \oplus (\mathbf{x})$, we then have that $((-1) \odot \mathbf{x}) \oplus (\mathbf{x}) = \mathbf{0}$.

We thus conclude that $((-1) \odot \mathbf{x})$ is an additive inverse of $\mathbf{x}$. Since the additive inverse of a vector space is unique, we have $((-1) \odot \mathbf{x}) = (-\mathbf{x})$. $\blacksquare$

Lemma 5: The cancellation identity

Let $\mathbf{x} \in V$, a vector space, and $a \in \mathbb{F}$.

$$\text{If } a \odot \mathbf{x} = \mathbf{0}, \text{ then either } a = 0 \text{ or } \mathbf{x} = \mathbf{0}.$$

Proof

Either $a = 0$, in which case the result is proven, or, $a \neq 0$, in which case then a^{-1} exists and by axiom (g), $(a^{-1}a) \odot \mathbf{x} = a^{-1} \odot (a \odot \mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$, using Lemma 3. $\blacksquare$

These results are simple and straightforwards, and we naturally use them in our day to day manipulations and calculations within any vector space.

Last time we introduced both the concepts of Span and Vector subspace.
It should be no surprise that the easiest way which we have to produce vectors subspaces is through spanning a subset of a vector space.

Lemma 6

Let V be a vector space, and let $W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\} \subset V$, where $p \geq 1$. Then

$Span(W)$ is a vector subspace of V and is the smallest subspace of V that contains W.

Proof

Let $W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p\} \subset V$, where $p \geq 1$.

Since $p \geq 1$, then $\mathbf{w}_1 \in W \subset V$, and so $\mathbf{w}_1 \in \mathit{Span}(W)$. Thus, $\mathit{Span}(W)$ is a non-empty subset of V.

If $\mathbf{x}$ and $\mathbf{y} \in \mathit{Span}(W)$, then there exist constants $a_i, b_i \in \mathbb{F}$, $i = 1, \ldots, p$, such that

$$\mathbf{x} = (a_1 \odot \mathbf{w}_1) \oplus (a_2 \odot \mathbf{w}_2) \oplus \cdots \oplus (a_p \odot \mathbf{w}_p) \quad \text{and}$$

$$\mathbf{y} = (b_1 \odot \mathbf{w}_1) \oplus (b_2 \odot \mathbf{w}_2) \oplus \cdots \oplus (b_p \odot \mathbf{w}_p),$$

and so

(i) $\mathbf{x} \oplus \mathbf{y} = (a_1 \odot \mathbf{w}_1) \oplus (a_2 \odot \mathbf{w}_2) \oplus ... \oplus (a_p \odot \mathbf{w}_p) \oplus (b_1 \odot \mathbf{w}_1) \oplus (b_2 \odot \mathbf{w}_2) \oplus ... \oplus (b_p \odot \mathbf{w}_p)$

$$= ((a_1 + b_1) \odot \mathbf{w}_1) \oplus ((a_2 + b_2) \odot \mathbf{w}_2) \oplus \cdots \oplus ((a_p + b_p) \odot \mathbf{w}_p).$$

We then conclude that $(\mathbf{x} \oplus \mathbf{y}) \in \mathit{Span}(W)$.

In addition

(ii) $c \odot \mathbf{x} = c \odot ((a_1 \odot \mathbf{w}_1) \oplus (a_2 \odot \mathbf{w}_2) \oplus \cdots \oplus (a_p \odot \mathbf{w}_p))$

$$= ((c \times a_1) \odot \mathbf{w}_1) \oplus ((c \times a_2) \odot \mathbf{w}_2) \oplus \cdots \oplus ((c \times a_p) \odot \mathbf{w}_p).$$

We then conclude that $(c \odot \mathbf{x}) \in \mathit{Span}(W)$.

Let U be another vector subspace of V which contains W. Then by closure, U must contain all linear combinations of elements of W, that is $\mathit{Span}(W) \subseteq U$.

We can conclude that $\mathit{Span}(W)$ is the smallest subspace of V that contains W. ■

Example 7: Vector subspaces

$$\mathit{Span}(\{1 - 2x + 3x^2 \, , \, 9 - 5x + x^2\})$$

$$= \{c_1(1 - 2x + 3x^2) + c_2(9 - 5x + x^2) : c_1, c_2 \in \mathbb{R}\}\,,$$

is a vector subspace of $P_2(\mathbb{R})$.

$$Span\left(\left\{\begin{pmatrix}1&3\\2&2\\3&1\end{pmatrix},\begin{pmatrix}-5&9\\-7&7\\-9&5\end{pmatrix},\begin{pmatrix}1&-1\\-1&1\\1&-1\end{pmatrix}\right\}\right)$$

$$=\left\{c_1\begin{pmatrix}1&3\\2&2\\3&1\end{pmatrix}+c_2\begin{pmatrix}-5&9\\-7&7\\-9&5\end{pmatrix}+c_3\begin{pmatrix}1&-1\\-1&1\\1&-1\end{pmatrix}:c_1,c_2,c_3\in\mathbb{R}\right\},$$

is a vector subspace of $M_{3\times 2}(\mathbb{R})$.

The next subject for discussion is that of linear dependence and independence.

Definition 5: Linear independence and dependence

Let V be a vector space and let $W=\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_p\}\subset V$.

We say that W is **linearly dependent** to mean that there exist scalars $a_i\in\mathbb{F}$, $i=1,...,p$,

not all zero, such that

$$(a_1\odot\mathbf{w}_1)\oplus(a_2\odot\mathbf{w}_2)\oplus\cdots\oplus(a_p\odot\mathbf{w}_p)=\mathbf{0}.$$

That is, the subset W, is linearly dependent when it is possible to construct the zero vector out of a non-trivial linear combination of vectors in W.

The set W is **linearly independent** when it is not linearly dependent.

So W is linearly independent means that the only solution to the equation

$$(a_1\odot\mathbf{w}_1)\oplus(a_2\odot\mathbf{w}_2)\oplus\cdots\oplus(a_p\odot\mathbf{w}_p)=\mathbf{0}$$

is the trivial solution, that is, if $a_1=0,\ a_2=0,\ldots,\ a_p=0$.

Example 8

Is the set $\left\{\begin{pmatrix}1&2\\3&4\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix},\begin{pmatrix}0&-1\\-1&0\end{pmatrix},\begin{pmatrix}4&3\\2&1\end{pmatrix},\begin{pmatrix}1&1\\1&1\end{pmatrix}\right\}$

linearly independent?

You may assume that we are using the usual addition and scalar multiplication in $M_{2\times 2}(\mathbb{R})$.

Solution

We consider the equation:

$$a\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + b\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + c\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} + d\begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix} + e\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

which is

$$\begin{pmatrix} a+b+4d+e & 2a-c+3d+e \\ 3a-c+2d+e & 4a+b+d+e \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

This yields a linear system of 4 equations in 5 unknowns.

The augmented matrix is:

$$\begin{pmatrix} 1 & 1 & 0 & 4 & 1 & | \, 0 \\ 2 & 0 & -1 & 3 & 1 & | \, 0 \\ 3 & 1 & 0 & 2 & 1 & | \, 0 \\ 4 & 0 & -1 & 1 & 1 & | \, 0 \end{pmatrix},$$

which row reduces to,

$$\begin{pmatrix} 1 & 0 & 0 & -1 & 0 & | \, 0 \\ 0 & 1 & 0 & 5 & 1 & | \, 0 \\ 0 & 0 & 1 & -5 & -1 & | \, 0 \\ 0 & 0 & 0 & 0 & 0 & | \, 0 \end{pmatrix}.$$

Since there are only three pivots, then there will be $5 - 3 = 2$ parameters in the solution set of the system of equations.

Thus there are many non-trivial solutions, and the set of matrices is linearly dependent.

Note that we knew before doing any calculations that this set is linearly dependent, as there are 4 equations for the 5 unknowns in a homogeneous system. Thus the rank of the coefficient matrix is at most 4 (it is actually 3) and we are guaranteed non-trivial solutions.

We can also extend some of our work in $\mathbb{F}^n$ to state that even though the set of 5 matrices is linearly dependent, we can choose from this set, a subset of 3 matrices which is both linearly independent and which has the same span as the five, for example the first three matrices.

We now combine the ideas of spanning and linear independence to construct a basis.

Definition 6

Let V be a vector space, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} \subset V$.

We say that B is a **basis** for V to mean that B is linearly independent and that $Span(B) = V$.

Thus B is a basis for V when

(i) $B \subset V$,

(ii) $Span(B) = V$

(iii) B is linearly independent.

Many of the vector spaces that we are used to have a natural standard basis associated with them.

Example 9

$\{1, x, x^2, \ldots, x^n\}$ is the standard basis for $P_n(\mathbb{F})$.

$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ is the standard basis for $M_{2\times 2}(\mathbb{F})$.

A basis is important due to the unique representation theorem, which we do not state again.

A basis both allows us to write our vectors down, and it allows us to do so in a unique manner.

Definition 7: Components and coordinates

Let V be a vector space and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V.

If $\mathbf{v} \in V$, then there exist unique scalars $a_1, a_2, \ldots, a_n$ such that

$$\mathbf{v} = (a_1 \odot \mathbf{v}_1) \oplus (a_2 \odot \mathbf{v}_2) \oplus \cdots \oplus (a_n \odot \mathbf{v}_n).$$

The scalars are referred to as the **coordinates**, or the **components** of the vector $\mathbf{v}$ in the basis B.

We use the notation $[\mathbf{v}]_B = (a_1, a_2, .., a_n)^T$, which is known as the **coordinate or the component vector of v** in the basis B.

Example 10

In $P_3(\mathbb{R})$, if we use the standard basis $S = \{1, x, x^2, x^3\}$, then we have

$$\left[2 - 3x + 4x^2 - 5x^3\right]_B = \begin{pmatrix} 2 \\ -3 \\ 4 \\ -5 \end{pmatrix},$$

In $M_{2\times 2}(\mathbb{C})$, using the standard basis $B = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$, then

$$\begin{bmatrix} 2+i & 5 \\ 3-i & 2-i \end{bmatrix}_B = \begin{pmatrix} 2+i \\ 5 \\ 3-i \\ 2-i \end{pmatrix}.$$

There are many results that we have had in $\mathbb{F}^n$ and which hold in a general vector space. These include:

Let V be a vector space and let $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V. Then every basis for V has exactly n vectors in it. We refer to n as the dimension of the vector space, and write $\dim(V) = n$.

Note that $\dim(P_n(\mathbb{F})) = n + 1$.

If we have two bases for the same vector space, then we can introduce the change of basis matrix, which will relate components in the two bases.

That is, let $B_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, and $B_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ be bases for the same vector space V, then the change of basis matrix from B_1 to B_2, is denoted by ${}_{B_2}[I]_{B_1}$, and is obtained as follows:

$${}_{B_2}[I]_{B_1} = \left([\mathbf{v}_1]_{B_2}, [\mathbf{v}_2]_{B_2}, .., [\mathbf{v}_n]_{B_2}\right).$$

The change of basis matrix from B_2 to B_1, is denoted by ${}_{B_1}[I]_{B_2}$.

It may either be obtained from the expression

$${}_{B_1}[I]_{B_2} = \left([\mathbf{w}_1]_{B_1}, [\mathbf{w}_2]_{B_1}, \ldots, [\mathbf{w}_n]_{B_1}\right).$$

or from the fact that ${}_{B_1}[I]_{B_2} = \left({}_{B_2}[I]_{B_1}\right)^{-1}$.

Example 11

Find the change of bases matrices for $P_3(\mathbb{R})$, ${}_S[I]_{B_1}$ and ${}_{B_1}[I]_S$, when

$$B_1 = \left\{1 + x^2, 1 - x^2, x - x^3, x + x^3\right\}.$$

Make use of the latter to obtain $\left[1 + 2x - 3x^2 - 4x^3\right]_{B_1}$.

Solution

$${}_S[I]_{B_1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix},$$

So that,

$${}_{B_1}[I]_S = \left({}_S[I]_{B_1}\right)^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 \end{pmatrix},$$

and then

$$\left[1 + 2x - 3x^2 - 4x^3\right]_{B_1} = {}_{B_1}[I]_S \left[1 + 2x - 3x^2 - 4x^3\right]_S$$

$$= \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ -3 \\ -4 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 3 \\ -1 \end{pmatrix}.$$

Find the change of bases matrices for $M_{2\times 2}(\mathbb{R})$, ${}_S[I]_{B_1}$ and ${}_{B_1}[I]_S$, when

$$B_1 = \left\{\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} -1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 1 & -2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ -3 & 4 \end{pmatrix},\right\}.$$

Make use of the latter to obtain $\left[\begin{pmatrix} 2 & -4 \\ 6 & -8 \end{pmatrix},\right]_{B_1}$.

Solution

$${}_S[I]_{B_1} = \begin{pmatrix} 1 & -1 & 1 & 1 \\ 2 & 2 & -2 & 2 \\ 3 & 3 & 3 & -3 \\ 4 & 4 & 4 & 4 \end{pmatrix}.$$

So that

$$ {}_{B_1}[I]_S = ({}_S[I]_{B_1})^{-1} = \frac{1}{24}\begin{pmatrix} 12 & 6 & 4 & -3 \\ -12 & 0 & 0 & 3 \\ 0 & -6 & 0 & 3 \\ 0 & 0 & -4 & 3 \end{pmatrix}, $$

and then

$$ \left[\begin{pmatrix} 2 & -4 \\ 6 & -8 \end{pmatrix}\right]_{B_1} = {}_{B_1}[I]_S \left[\begin{pmatrix} 2 & -4 \\ 6 & -8 \end{pmatrix}\right]_S $$

$$ = \frac{1}{24}\begin{pmatrix} 12 & 6 & 4 & -3 \\ -12 & 0 & 0 & 3 \\ 0 & -6 & 0 & 3 \\ 0 & 0 & -4 & 3 \end{pmatrix}\begin{pmatrix} 2 \\ -4 \\ 6 \\ -8 \end{pmatrix} = 2\begin{pmatrix} 1 \\ -1 \\ 0 \\ -1 \end{pmatrix}. $$

Topic 22

The Rowspace of a Matrix

In this topic, we will consider the vector space $M_{1\times n}(\mathbb{F})$, the space of 1 by n matrices. This space is important because of the fact that if $A \in M_{m\times n}(\mathbb{F})$, then one way of thinking about A is that

$$A = \begin{pmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^m \end{pmatrix},$$

i.e., A is built up from m row vectors, $\mathbf{A}^1, \mathbf{A}^2, \ldots, \mathbf{A}^m$, each one of which is in $M_{1\times n}(\mathbb{F})$.

Note the two vector spaces $M_{1\times n}(\mathbb{F})$ and $M_{n\times 1}(\mathbb{F})$, are not the same.
We have considered from the beginning of this course, the vector space $M_{n\times 1}(\mathbb{F})$ of column vectors, or just vectors, as we have identified them with $\mathbb{F}^n$, or to be more precise, we have identified them with the components of vectors in $\mathbb{F}^n$, once a basis has been declared. These are column vectors. The vectors in $M_{1\times n}(\mathbb{F})$ are rows of numbers, not columns of numbers, and it is incorrect to claim that they are the same: e.g.

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \neq (1, 2, 3)\,.$$

There is, however, a very natural isomorphism (invertible linear transformation) between the two spaces: the transpose operation.
We have used this often as a space saving device, e.g.

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = (1, 2, 3)^T\,.$$

Definition 1: Rowspace

Let $A \in M_{m\times n}(\mathbb{F})$. We define the **rowspace** of A, denoted $Row(A)$, by

$$Row(A) = Span(\{\mathbf{A}^1, \mathbf{A}^2, \ldots, \mathbf{A}^m\}).$$

Notice that $Row(A)$ is a vector subspace of $M_{1\times n}(\mathbb{F})$.

Example 1

$$A = \begin{pmatrix} 1 & 4 & 1 \\ 2 & 3 & -1 \\ 3 & 2 & 1 \\ 4 & 1 & -1 \end{pmatrix},$$

and so

$$Row(A) = Span(\{(1,4,1),(2,3,-1),(3,2,1),(4,1,-1)\}).$$

We are interested in the following standard linear algebra questions about $Row(A)$.

i) How big is $Row(A)$, i.e. what is $\dim(Row(A))$?

ii) What is a basis for $Row(A)$, B_1, and how do we obtain it efficiently?

iii) How do we extend B_1 to a basis for $M_{1\times n}(\mathbb{F})$?

Lemma 1

Let $A \in M_{m\times n}(\mathbb{F})$. Suppose that B is obtained by performing a finite sequence of elementary row operations on A, then

$$Row(B) = Row(A).$$

Proof

We begin by noting that each row of B is a linear combination of the rows of A, and since elementary row operations are invertible, each row of A is a linear combination of the rows of B. We have

$$\mathbf{B}^i = \sum_{j=1}^{m} \alpha_{ij}\,\mathbf{A}^j, \quad \text{for some constants } \alpha_{ij}, \quad \text{for each } i = 1, \ldots, m,$$

and so,

$$\text{if } \mathbf{x} \in Row(\mathbf{B}), \quad \text{then } \mathbf{x} = \sum_{k=1}^{m} c_k\,\mathbf{B}^k \;\; \text{for some constants } c_k,\; k = 1, \ldots, m.$$

It follows that

$$\mathbf{x} = \sum_{k=1}^{m}\sum_{j=1}^{m} \alpha_{kj} c_k \mathbf{A}^j, \quad \text{and so } \mathbf{x} \in Row(\mathbf{A}).$$

If $\mathbf{y} \in Row(\mathbf{A})$, then a similar proof shows that $\mathbf{y} \in Row(\mathbf{B})$.
We conclude that the two spaces are identical, that is

$$Row(B) = Row(A). \qquad \blacksquare$$

Corollary 1

$dim(Row(A)) = rank(A)$.

Proof

Suppose we perform a finite number of elementary row operations on A and arrive at B, a row echelon from of A. There are exactly r non-zero rows in B, these rows are linearly independent, as can be seen by the positions of the pivots, and, in particular, that there are zeros in all the rows lying below any pivot. Thus the rows of B that have pivots in them form a basis for $Row(B)$, and for $Row(A)$. It now follows that

$$r = rank(A) = rank(B) = \text{the number of vectors in a basis for } Row(A),$$

that is,

$$dim(Row(A)) = rank(A). \qquad \blacksquare$$

Warning: the rows of A corresponding to the rows in B which have pivots in them, i.e. the first r rows of A, are not necessarily linearly independent. It is the first r rows of B that are linearly independent.

Example 2

Find a basis for $Row(A)$, and extend it to a basis of $M_{1\times 5}$. Then find a basis for $Row(A)$, which uses the rows of A, and extend it to a basis of $M_{1\times 5}$, if

$$A = \begin{pmatrix} 1 & 1 & 5 & 0 & 1 \\ 2 & 0 & 4 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 3 & 1 & 3 & 0 & 1 \\ 4 & 0 & 2 & 1 & 1 \\ 5 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Solution

We begin by row reducing this matrix to get:

$$B = \begin{pmatrix} 1 & 0 & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Note that the choice was made to reduce this matrix to reduced row echelon form.

We can read off from this that $rank(A) = 3$, and a basis for $Row(A)$ is

$$B_1 = \left\{ \left(1, 0, 0, \frac{1}{6}, \frac{1}{6}\right), (0, 1, 0, -1, 0), \left(0, 0, 1, \frac{1}{6}, \frac{1}{6}\right) \right\} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}.$$

In order to obtain a basis for $M_{1\times5}$, which contains these three vectors, we add the standard basis for $M_{1\times5}$ to this set and then we reduce it.

Let $U = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6, \mathbf{v}_7, \mathbf{v}_8\}$ with $\mathbf{v}_4 = (1, 0, 0, 0, 0), \ldots, \mathbf{v}_8 = (0, 0, 0, 0, 1)$.

We consider the equation:

$$c_1\mathbf{v}_1 + \cdots + c_8\mathbf{v}_8 = \mathbf{0}, \quad \text{which has coefficient matrix:}$$

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{6} & -1 & \frac{1}{6} & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We reduce this coefficient matrix to get:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

From which we conclude that a basis for $M_{1\times5}$ which contains B_1, is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_5, \mathbf{v}_6\}$.

If we now want a basis for $Row(A)$ using rows of A, then we have to re-examine the matrix A. We know that we need to obtain three linearly independent row vectors from A, it is also clear that the third row of A will be of no use to us.

We consider the system of equations:

$$c_1\mathbf{A}^1 + \cdots + c_6\mathbf{A}^6 = \mathbf{0}, \quad \text{for which the coefficient matrix is:}$$

$$\begin{pmatrix} 1 & 2 & 0 & 3 & 4 & 5 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 5 & 4 & 0 & 3 & 2 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Notice that this is the transpose of the original matrix A. Reducing it, we get

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

Thus rows 1, 2 and 4 of A are linearly independent, and may be used to obtain a basis for $Row(A)$. That is, a basis for the rowspace of A using the rows of A is

$$B_2 = \{(1,1,5,0,1),(2,0,4,1,1),(3,1,3,0,1)\} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}.$$

If we want to obtain a basis for $M_{1\times 5}$, which contains these three vectors, we add the standard basis for $M_{1\times 5}$ to this set and then we reduce it.

Let $U = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5, \mathbf{u}_6, \mathbf{u}_7, \mathbf{u}_8\}$ with $\mathbf{u}_4 = (1,0,0,0,0), \ldots, \mathbf{u}_8 = (0,0,0,0,1)$.

We consider the equation:

$$c_1\mathbf{u}_1 + \cdots + c_8\mathbf{u}_8 = \mathbf{0}, \quad \text{which has coefficient matrix:}$$

$$\begin{pmatrix} 1 & 2 & 3 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 5 & 4 & 3 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

This reduces to

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{3}{2} \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -1 \end{pmatrix}.$$

We conclude that rows, one, two and four of A together with the first two vectors of the standard basis, form a basis for $M_{1\times 5}$.

Lemma 2

Let $A \in M_{m\times n}(\mathbb{F})$. Then $rank(A) = rank(A^T)$.

Proof

In order to obtain the rank of A, we could row reduce A into echelon form and the number of pivots is equal to the rank of A. This number also tells us the number of linearly independent rows that A has. The rows of A are the columns of A^T. Thus the rank of A also gives us the number of linearly independent columns of A^T, that is, the rank of A^T. $\blacksquare$

TOPIC 23
Matrix Representations of Linear Transformations

In our consideration of linear transformations, we have spent most of the time dealing with linear operators, for which the domain and codomain are both $\mathbb{F}^n$. There are some very good reasons for this: namely that the matrix representation of the linear operator, in any basis, is a square matrix, and so we can examine issues such as the eigenvalue, problem, diagonalizability and invertibility.

We now return briefly to discuss linear transformations whose domains and codomains are not necessarily identical.

Definition 1: Linear transformation

Let $T : U \to V$ be a function from a subspace U of $\mathbb{F}^n$ to a subspace V of $\mathbb{F}^m$
We then say that T is a **linear transformation** from U to V, to mean that

a) $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$, $\forall \mathbf{u}_1, \mathbf{u}_2 \in U$, **and**

b) $T(c\mathbf{u}_1) = cT(\mathbf{u}_1), \forall \mathbf{u}_1 \in U$ and $\forall c \in \mathbb{F}$.

Definition 2: Matrix representation of a linear transformation

Let $T : U \to V$ be a linear transformation from U to V.

Let $B_1 = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ be a basis for U and $B_2 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_q\}$ be a basis for V.

We then define the matrix representation of the linear transformation (with respect to B_1 of U and B_2 of V), ${}_{B_2}[T]_{B_1}$, to be the $(q \times p)$ matrix obtained as follows:

$${}_{B_2}[T]_{B_1} = \left([T(\mathbf{u}_1)]_{B_2}, [T(\mathbf{u}_2)]_{B_2}, \ldots, [T(\mathbf{u}_p)]_{B_2}\right).$$

That is, as usual, we find the images of the basis vectors in the domain, and write them, as column vectors, in term of the basis vectors in the codomain. Since the domain and codomain are not necessarily the same and/or may need not necessarily use the same basis in both the domain and codomain, we have to specify both of these bases.

In the simplest case, we use the standard basis in the domain S_1, and the standard basis in the codomain S_2. We would then have

$${}_{S_2}[T]_{S_1} = \left([\,T(\mathbf{e}_1)\,]_{S_2}, [T(\mathbf{e}_2)]_{S_2}, \ldots, [T(\mathbf{e}_p)]_{S_2}\right),$$

which we would normally write as

$$ {}_{S_2}[T]_{S_1} = (T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_p)) \,. $$

We have just been writing

$$ {}_{S_2}[T]_{S_1} = [T]_S = (T(\mathbf{e}_1), T(\mathbf{e}_2), \ldots, T(\mathbf{e}_p)) \,. $$

The simplest example of a linear transformation arises when we have a matrix: that is,

if $A \in M_{p\times q}(\mathbb{F})$, then $T_A : \mathbb{F}^q \to \mathbb{F}^p$ is defined by $T_A(\mathbf{x}) = A\mathbf{x}$, $\forall \mathbf{x} \in \mathbb{F}^q$, and $[T_A]_S = A$.

Lemma 1

Let $T : U \to V$ be a linear transformation from U to V, with basis B_1 for U and basis B_2 for V.

Let ${}_{B_2}[T]_{B_1}$ be the matrix representation of the linear transformation. Then

$$ [T(\mathbf{x})]_{B_2} = {}_{B_2}[T]_{B_1}\, [\mathbf{x}]_{B_1}, \ \forall \mathbf{x} \in U. $$

We have seen similar results for the case of a linear operator.

Example 1

Let T be the linear transformation given by $T : \mathbb{C}^2 \to \mathbb{C}^3$, such that

$$ T\begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} z+w \\ -z+w \\ 2z+3w \end{pmatrix}. $$

Find $[T]_S$ and use it to evaluate $T\begin{pmatrix} 1+i \\ 1-i \end{pmatrix}$.

Solution

$$ T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \quad \text{and} \quad T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}. $$

Thus, $\qquad [T]_S = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 3 \end{pmatrix}.$

We then have

$$\left[T\begin{pmatrix}1+i\\1-i\end{pmatrix}\right]_S = [T]_S\left[\begin{pmatrix}1+i\\1-i\end{pmatrix}\right]_S$$

$$= \begin{pmatrix}1 & 1\\-1 & 1\\2 & 3\end{pmatrix}\begin{pmatrix}1+i\\1-i\end{pmatrix} = \begin{pmatrix}2\\-2i\\5-i\end{pmatrix}.$$

It is of course possible to make use of a different basis, B_3 in the domain, and a different basis, B_4, in the codomain. We would then have ${}_{B_4}[T]_{B_3}$.

It will come as no surprise that the two matrix representations, ${}_{B_2}[T]_{B_1}$ and ${}_{B_4}[T]_{B_3}$, are related to each other, in the natural way, through two change of basis matrices, one in the domain and one in the codomain:

$${}_{B_4}[T]_{B_3} = {}_{B_4}[I]_{B_2}\ {}_{B_2}[T]_{B_1}\ {}_{B_1}[I]_{B_3}.$$

We then have:

$$\begin{aligned}[T(\mathbf{x})]_{B_4} &= {}_{B_4}[T]_{B_3}\ [\mathbf{x}]_{B_3}\\ &= {}_{B_4}[I]_{B_2}\ {}_{B_2}[T]_{B_1}\ {}_{B_1}[I]_{B_3}\ [\mathbf{x}]_{B_3}\\ &= {}_{B_4}[I]_{B_2}\ {}_{B_2}[T]_{B_1}\ [\mathbf{x}]_{B_1}\\ &= {}_{B_4}[I]_{B_2}\ [T(\mathbf{x})]_{B_2}\end{aligned}$$

Example 2

Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is such that ${}_{S_2}[T]_{S_1} = \begin{pmatrix}1 & 4\\2 & 5\\3 & 6\end{pmatrix}$.

We are given that $B_1 = \left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right\}$ is a basis for $\mathbb{R}^2$ and

$B_2 = \left\{\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\-1\end{pmatrix}\right\}$ is a basis for $\mathbb{R}^3$.

Evaluate ${}_{B_2}[T]_{B_1}$.

Solution

We first calculate :

$$ {}_{S_1}[I]_{B_1} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \text{and} \quad {}_{S_2}[I]_{B_2} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}, $$

so that

$$ {}_{B_2}[I]_{S_2} = ({}_{S_2}[I]_{B_2})^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}^{-1} = \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}. $$

We now now obtain ${}_{B_2}[T]_{B_1}$:

$$ \begin{aligned} {}_{B_2}[T]_{B_1} &= {}_{B_2}[I]_{S_2}\ {}_{S_2}[T]_{S_1}\ {}_{S_1}[I]_{B_1} \\ &= \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 7 & -3 \\ 0 & 0 \\ -2 & 0 \end{pmatrix}. \end{aligned} $$

Example 3

Find the matrix representation for the linear transformation T, projection on the plane P given by $2x - 5y + 8z = 0$.

We consider T as a function from $\mathbb{R}^3$ to P. We will evaluate ${}_{B}[T]_{S}$, where B is a basis chosen to be adapted to the plane, but will make use of an intermediate basis B_1 for $\mathbb{R}^3$.

Solution

We let $B = \left\{ \begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix} \right\}$ and $B_1 = \left\{ \begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -5 \\ 8 \end{pmatrix} \right\}$.

We then have

$$ T\left(\begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix} = 1\begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix} + 0\begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix} $$

$$T\left(\begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix} = 0\begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix}$$

$$T\left(\begin{pmatrix} 2 \\ -5 \\ 8 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix} + 0\begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix}$$

We thus have that ${}_B[T]_{B_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

In order to obtain ${}_B[T]_S$, we will need ${}_{B_1}[I]_S$. We know that

$${}_S[I]_{B_1} = \begin{pmatrix} 5 & 4 & 2 \\ 2 & 0 & -5 \\ 0 & -1 & 8 \end{pmatrix}, \quad \text{so that}$$

$${}_{B_1}[I]_S = (\,{}_S[I]_{B_1}\,)^{-1} = \frac{1}{93}\begin{pmatrix} 5 & 34 & 20 \\ 16 & -40 & -29 \\ 2 & -5 & 8 \end{pmatrix}.$$

We thus have

$${}_B[T]_S = {}_B[T]_{B_1}\ {}_{B_1}[I]_S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\frac{1}{93}\begin{pmatrix} 5 & 34 & 20 \\ 16 & -40 & -29 \\ 2 & -5 & 8 \end{pmatrix}$$

$$= \frac{1}{93}\begin{pmatrix} 5 & 34 & 20 \\ 16 & -40 & -29 \end{pmatrix}.$$

For example,

$$\left[T\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}\right]_B = \frac{1}{93}\begin{pmatrix} 5 & 34 & 20 \\ 16 & -40 & -29 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{93}\begin{pmatrix} 133 \\ -151 \end{pmatrix}, \text{ meaning that}$$

$$T\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{133}{93}\begin{pmatrix} 5 \\ 2 \\ 0 \end{pmatrix} - \frac{151}{93}\begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix} = \frac{1}{93}\begin{pmatrix} 61 \\ 266 \\ 151 \end{pmatrix}.$$

Exercise: Consider the linear transformation, L, projection onto the plane $2x - 5y + 8z = 0$ as a function from $\mathbb{R}^3$ to $\mathbb{R}^3$. Find $[T]_S$ and verify the image of $(1, 2, 3)^T$.

Example 5

Consider $T_I : \mathbb{F}^n \to \mathbb{F}^n$ given by $T_I(\mathbf{x}) = \mathbf{x},\ \forall \mathbf{x} \in \mathbb{F}^n$. Clearly ${}_S[T_I]_S = I_n$.

Let us use different bases for the domain and for the codomain.

For the domain, we have $B_1 = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, and for the codomain we have $B_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$.

We then have :

$$\begin{aligned} {}_{B_2}[T_I]_{B_1} &= \big([T_I(\mathbf{u}_1)]_{B_2} \, , \, [T_I(\mathbf{u}_2)]_{B_2} \, , \ldots , \, [T_I(\mathbf{u}_n)]_{B_2} \big) \\ &= \big([\mathbf{u}_1]_{B_2} \, , \, [\mathbf{u}_2]_{B_2} \, , \ldots , \, [\mathbf{u}_n]_{B_2} \big) \\ &= {}_{B_2}[I]_{B_1} , \end{aligned}$$

and this is an old friend, the change of basis matrix from B_1 coordinates to B_2 coordinates. This is the reason why we use that specific notation for the change of basis matrix.

You will learn more about trying to simplify non-square matrices when you study the concept of singular value decomposition in MATH 235.

Topic 24

The Invertible Matrix Theorem

During our procession through the course material, the concept of invertibility has arisen at various points and in various disguises. We bring (many) of them together in a result known as **The Invertible Matrix Theorem**.

Theorem 1: The Invertible Matrix Theorem

Let $A \in M_{n\times n}(\mathbb{F})$. Then the following statements are equivalent.

(1) A is invertible.

(2) $\exists B \in M_{n\times n}(\mathbb{F}) \, : \, AB = BA = I_n$. Definition 10 in Topic 13C.

(3) $A\mathbf{x} = \mathbf{b}$ has a unique solution $\forall \mathbf{b} \in \mathbb{F}^n$. Lemma 16 and Corollary 5 in Topic 13C.

(4) T_A is an invertible linear transformation. Lemma 16in Topic 13C.

(5) $[T_A]_B$ is invertible $\forall$ basis B of $\mathbb{F}^n$. Lemma 1 in Topic 16B and Lemma 2 in Topic 18.

(6) T_A is both one-to-one and onto. Definition 12 in Topic 13C.

(7) T_A is one-to-one. Remark 6 in Topic 13C

(8) T_A is onto. Remark 6 in Topic 13C.

(9) $rank(A) = n$. Lemma 2 in Topic 14.

(10) A has n pivots. Definition 1 in Topic 9.

(11) $dim(Col(A)) = n$. Remark 1 in Topic 22.

(12) $Col(A) = \mathbb{F}^n$. Corollary 1 in Topic 13A and statement 8.

(13) The columns of A are linearly independent. Lemma 5 in Topic 17B.

(14) The columns of A form a basis for $\mathbb{F}^{n.}$ Statement (13).

(15) $nullity(A) = 0$. Corollary 3 in Topic 13A.

(16) $N(A) = \{\mathbf{0}\}$. Corollary 3 in Topic 13A and statement 9.

(17) The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Statement (16).

(18) $\exists B \in M_{n\times n}(\mathbb{F}) : AB = I_n$. Lemma 1 in Topic 14.

(19) $\exists C \in M_{n\times n}(\mathbb{F}) : CA = I_n$. Similar to Lemma 1 in Topic 14.

(20) $RREF(A) = I_n$. Remark 1 in Topic 14.

(21) A may be written as the product of elementary matrices.
Lemma 14 in Topic 13C and proof of Lemma 3 in Topic 14.

(22) $\det(A) \neq 0$. Corollary 7 in Topic 15B.

(23) 0 is not an eigenvalue of A. Corollary 1 in Topic 16B .

(24) 0 is not a root of $\Delta_A(t)$. Definition 3 Topic 16A.

(25) $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b}$ in $\mathbb{F}^n$. Topic 14 Lemma 1.

(26) $dim(Row(A)) = n$. Corollary 1 in Topic 22.

(27) The row vectors of A are linearly independent. (26)

(28) The row vectors of A span $M_{1\times n}(\mathbb{F})$. (26)

(29) $N(T_A) = \{\mathbf{0}\}$.

(30) A^T is invertible. Topic 22 Lemma 2.

(31) $P^{-1}AP$ is invertible, where P is any invertible $M_{n\times n}(\mathbb{F})$ matrix.