School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 1

CLOSURE

Milan Pahor

MATH1231 ALGEBRA

CLOSURE

A subset S of a vector space V is said to be closed under addition if

$$\mathbf{v}_1, \mathbf{v}_2 \in S \implies \mathbf{v}_1 + \mathbf{v}_2 \in S.$$

A subset S of a vector space V is said to be closed under scalar multiplication if

$$\mathbf{v}_1 \in S \text{ and } \alpha \text{ a scalar } \implies \alpha\mathbf{v}_1 \in S.$$

LECTURE 1

CLOSURE IN $\mathbb{R}^2$ and $\mathbb{R}^3$

There are many ways of perceiving the concept of a vector. In Physics, a vector is a quantity with magnitude and direction, represented using arrows.

In Math1131, vectors were treated as algebraic objects, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}$ is a vector in three dimensional space $\mathbb{R}^3$.

In this course we will develop a completely abstract theory of vectors. You will be surprised to learn that every mathematical object you have ever used, functions, matrices, polynomials,........., everything, can be viewed as a vector if you so wish.

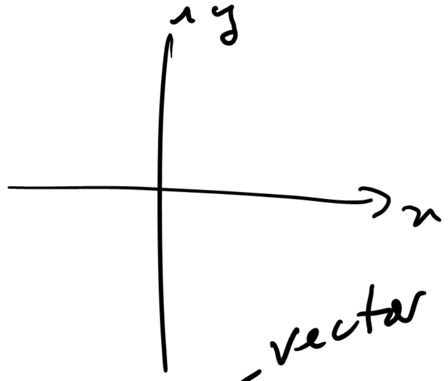
So what is a vector now? A vector is quite simply an element of a vector space! What is a vector space then?? That is a deep question which will be addressed soon.

In this lecture we will start the process by looking at the central concept of **closure** in the two dimensional vector space $\mathbb{R}^2$ and the three dimensional vector space $\mathbb{R}^3$. We will start with these two concrete spaces as you are familiar with them, and it is a little easier to see what is going on. Indeed throughout the course we will retreat to $\mathbb{R}^2$ and $\mathbb{R}^3$ when we need to develop a feeling for our vector space definitions and theorems.

But all of our concepts will always operate in a much wider setting, and we will discuss abstract vector spaces in Lecture 2.

The nice thing about $\mathbb{R}^2$ and $\mathbb{R}^3$ is that vector addition and scalar multiplication is trivial in these spaces, and we have simple pictures and definitions to help us perceive the objects under consideration:

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \; x, y \in \mathbb{R} \right\}$$



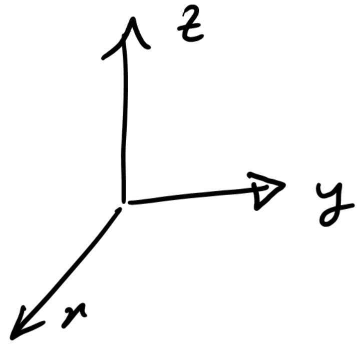
vector +

$$\begin{pmatrix} 1 \\ 4 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$$

$$7 \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 7 \\ 28 \end{pmatrix}$$

scalar ×

$$\mathbb{R}^3 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \; x, y, z \in \mathbb{R} \right\}$$



$$\begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 8 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 15 \end{pmatrix}$$

$$2 \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \\ 14 \end{pmatrix}$$

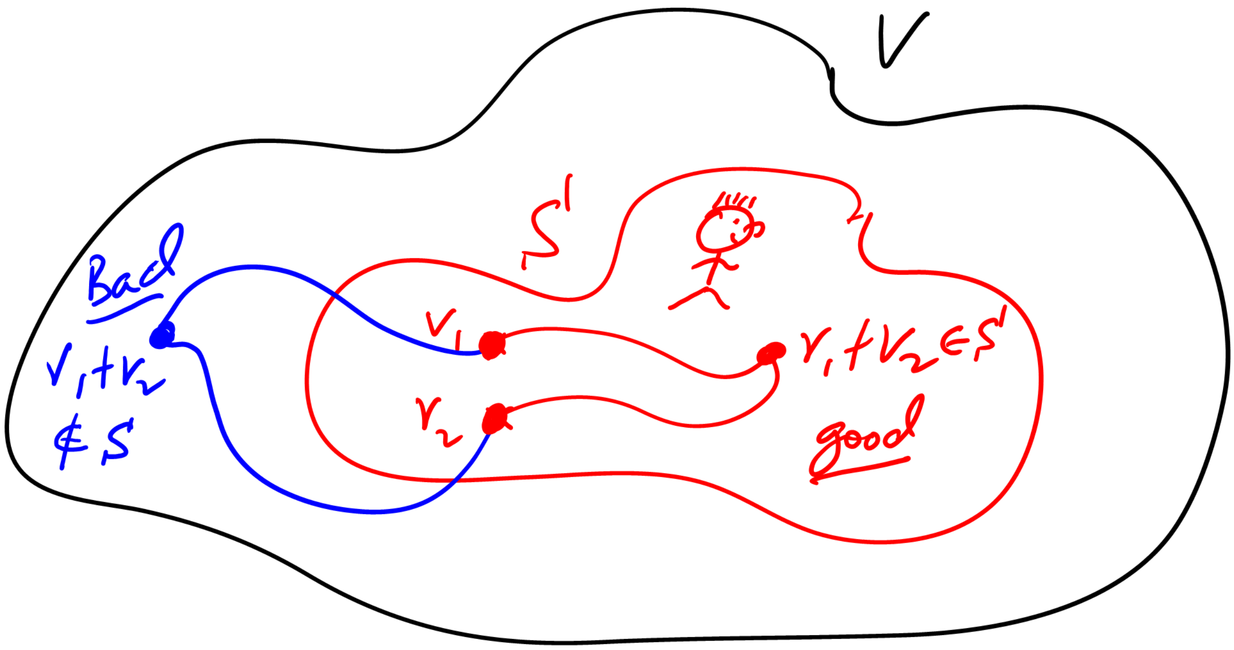
★

Our first task is to find within these spaces, smaller subsets which behave properly in their own right. Little baby spaces called subspaces which mimic the essential structure of the parent spaces. There are two properties we like to see, **closure under addition** and **closure under scalar multiplication**.

Definition: A subset S of a vector space V is said to be closed under addition if

$$\mathbf{v}_1, \mathbf{v}_2 \in S \implies \mathbf{v}_1 + \mathbf{v}_2 \in S$$

Discussion: What does this mean??



When two vectors belonging to S are added the vector sum MUST also be in S.

Example 1: Let S be the subset of $\mathbb{R}^3$ defined by

$$S = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; z = x^2 + y^2 \right\}$$

Discussion: What does this mean??

<u>Not</u> any old $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is in S

The z coordinates must be the sum of the squares of x & y coordinates.

a) Write down two non-trivial vectors in S. non-zero

b) Prove that S is **NOT** closed under addition.

a) $\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 25 \end{pmatrix} \in S$ ✓ since $5 = 1^2 + 2^2$ and $25 = 9 + 16$.

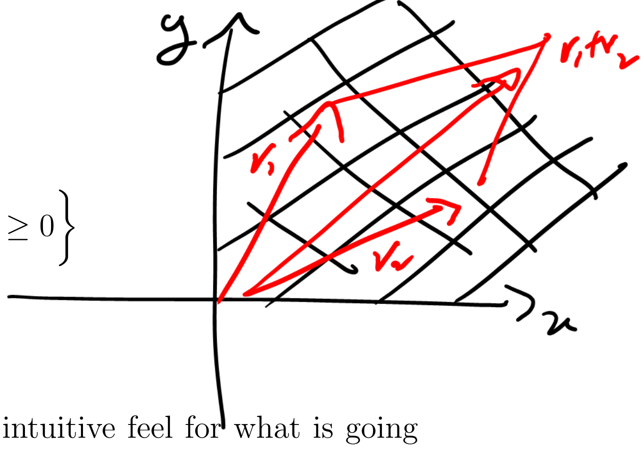
b) $\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} + \begin{pmatrix} 3 \\ 4 \\ 25 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \\ 30 \end{pmatrix} \notin S$ ✗

Since $30 \neq 4^2 + 6^2$

★

Note that when proving failure of closure using particular vectors, it is best to choose "interesting" vectors which have the potential to cause trouble. Never choose a zero vector.

Example 2: Let S be the first quadrant in $\mathbb{R}^2$. That is



$$S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; x \geq 0 \text{ and } y \geq 0 \right\}$$

Prove that S is closed under addition.

Sketch: The nice thing about $\mathbb{R}^2$ is that we have an intuitive feel for what is going on, using tip to tail addition for geometric vectors.

It is easy to prove that a set is **not** closed under addition. All you have to produce is one specific example of things going wrong. But to prove that a set **is** closed under addition you must provide a completely general argument. **You cannot and must not deal with specific vectors!**

Proof: Look very carefully at the language here. Our approach is to always:

1) Choose two random vectors in S and explain what is special about these vectors.

2) Add the vectors.

3) Carefully prove that the vector sum is **also** special and hence in S.

1) Let $\underset{\sim}{v}_1, \underset{\sim}{v}_2 \in S$

$\underset{\sim}{v}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\underset{\sim}{v}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$

with $x_1, y_1 \geq 0$ and $x_2, y_2 \geq 0$

2) $\underset{\sim}{v}_1 + \underset{\sim}{v}_2 = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}$

Clearly $x_1 + x_2 \geq 0$
and $y_1 + y_2 \geq 0$

$\therefore$ $\underset{\sim}{v}_1 + \underset{\sim}{v}_2 \in S$

$\therefore$ S is closed under addition

★

So some subsets of a vector space are closed under addition.....some are not. We like the ones that are!

A second “nice” property is closure under scalar multiplication. For most of our spaces the scalars we use will be the field of real numbers. Thus for example we can triple a vector $\mathbf{v}$ to get $3\mathbf{v}$, flip it and halve it to produce $-\frac{1}{2}\mathbf{v}$, or even scale it by π to get $\pi\mathbf{v}$.

It is also possible to use complex numbers as scalars and we will sometimes (rarely) do this. If nothing special is mentioned you may assume that the scalars are real numbers.

Definition: A subset S of a vector space V is said to be closed under scalar multiplication if

$$\mathbf{v}_1 \in S \text{ and } \alpha \text{ a scalar} \quad \Longrightarrow \quad \alpha\mathbf{v}_1 \in S.$$

Example 3: Let S be, once again, the first quadrant in $\mathbb{R}^2$. That is

$$S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; x \geq 0 \text{ and } y \geq 0 \right\}$$

a) Write down a vector in S.

b) Prove that S is not closed under scalar multiplication.

a) $\begin{pmatrix} 3 \\ 4 \end{pmatrix} \in S$ since $3 \geq 0,\ 4 \geq 0$.

Let $\alpha = -7$ be a scalar.

$\alpha\begin{pmatrix} 3 \\ 4 \end{pmatrix} = -7\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} -21 \\ -28 \end{pmatrix} \notin S$

$\therefore$ S is <u>not</u> closed under scalar $\times$.

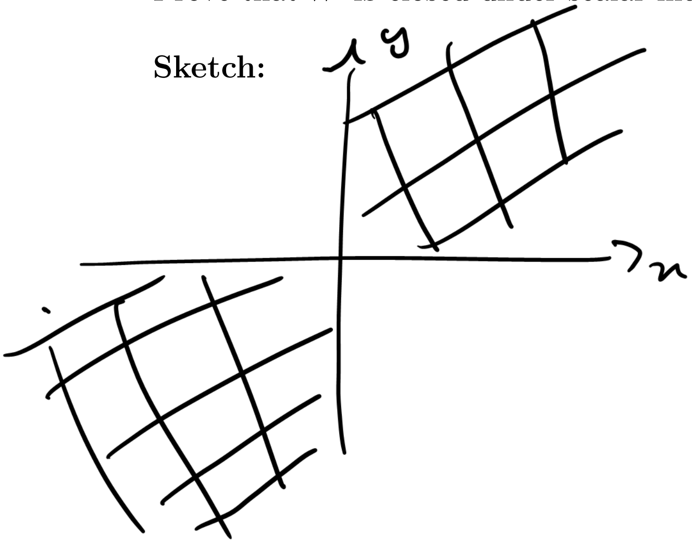
We can fix this!

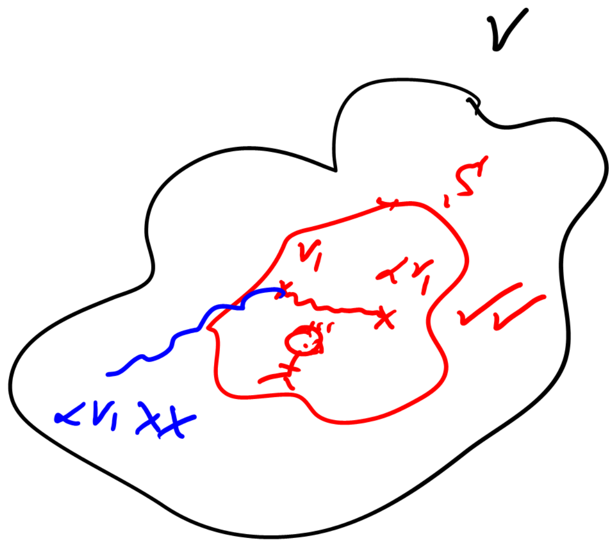
Example 4: Let W be the union of quadrants 1 and 3 in $\mathbb{R}^2$. That is

$$W = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; x \geq 0 \text{ and } y \geq 0 \quad \text{OR} \quad x \leq 0 \text{ and } y \leq 0 \right\}$$

Prove that W **is** closed under scalar multiplication.

Sketch:





Proof: Our approach for proofs of closure under scalar multiplication is to always:

1) Choose a random vector in S and explain what is special about this vector.

2) Multiply the vector by a random scalar α.

3) Carefully prove that the scaled vector is **also** special and hence in S.

1) So let $\mathbf{v}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \in W$.

Then either both x_1 and y_1 are greater than or equal to 0 OR both x_1 and y_1 are less than or equal to 0. That is, $\mathbf{v}_1$ is either in the first quadrant or the third quadrant.

2) Now let $\alpha \in \mathbb{R}$ be any scalar.

3) If $\alpha > 0$ then $\alpha\mathbf{v}_1$ will be in the same quadrant as $\mathbf{v}_1$ and thus be in W.

If $\alpha < 0$ then $\alpha\mathbf{v}_1$ will move $\mathbf{v}_1$ from quadrant 1 to 3 or vice versa. Either way, $\alpha\mathbf{v}_1$ will still be in W.

If $\alpha = 0$ then $\alpha\mathbf{v}_1 = \mathbf{0}$ which is certainly still in W.

Thus W is closed under scalar multiplication.

★

Discussion: Is W from Example 4 closed under vector addition? NO

Proof Let $\underset{\sim}{v}_1 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$, $\underset{\sim}{v}_2 = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$

Then $\underset{\sim}{v}_1, \underset{\sim}{v}_2 \in W$

$$\underset{\sim}{v}_1 + \underset{\sim}{v}_2 = \begin{pmatrix} 3 \\ 3 \end{pmatrix} + \begin{pmatrix} -2 \\ -4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \notin W$$

$\in$ Quadrant ★ 4 !!

So the above subset W is closed under scalar multiplication but **not** closed under vector addition. The opposite can also happen.

We prefer all of our spaces to be closed, since if they are not, simple processes such as vector addition and scalar multiplication are impossible to work with. If a space is not closed under vector addition, every time we added two vectors we would be terrified that the sum would literally disappear in our hands.

For this entire lecture we have played with $\mathbb{R}^2$ and $\mathbb{R}^3$. But what is a vector space in general? In the next lecture we will define abstract vector spaces and the subspaces which lurk within.

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 2

VECTOR SPACES

Milan Pahor

MATH1231 ALGEBRA
VECTOR SPACES

A vector space V is a set of mathematical objects which can be added and scaled "properly". Ten axioms need to be satisfied. The big three are:

Closure under vector addition: $\mathbf{v}_1, \mathbf{v}_2 \in V \implies \mathbf{v}_1 + \mathbf{v}_2 \in V$.

Closure under scalar multiplication: $\mathbf{v}_1 \in V$ and α a scalar $\implies \alpha\mathbf{v}_1 \in V$.

The existence of a zero vector $\mathbf{0}$.

In the previous lecture we looked at various subsets of $\mathbb{R}^2$ and $\mathbb{R}^3$, paying particular attention to whether or not the subsets were closed under vector addition and/or scalar multiplication. The spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ are examples of what we call vector spaces. In this lecture we will define vector spaces very carefully.

You need to come to grips with the fact that every mathematical object you have ever seen, polynomials, matrices, functions, complex numbers etc etc, can be perceived as a vector, living in a vector space. This vast generality is what gives the theory of vector spaces its immense power.

Roughly speaking, a vector space V is a collection of like-minded mathematical objects which can be added and scaled intelligently. That is, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are two elements of V, there must be some way of computing the vector sum $\mathbf{v}_1 + \mathbf{v}_2$ and some procedure for calculating a scalar multiple $\alpha\mathbf{v}_1$, where the scalar α is usually real.

Let's have a look at some vector spaces which are more exotic than just $\mathbb{R}^2$ and $\mathbb{R}^3$.

Example 1: For each of the following vector spaces V:

- Display the process of vector addition using a pair of non-trivial elements $\mathbf{v}_1$ and $\mathbf{v}_2$.
- Display the process of scalar multiplication using the scalar $\alpha = -5$.
- Write down the zero vector.

---

a) $V = \mathbb{R}^4$. This is the standard four dimensional real space.

$$\underbrace{\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}}_{\in \mathbb{R}^4} + \underbrace{\begin{pmatrix} 1 \\ 4 \\ 1 \\ 7 \end{pmatrix}}_{\in \mathbb{R}^4} = \begin{pmatrix} 2 \\ 6 \\ 4 \\ 11 \end{pmatrix}$$

$$-5 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} -5 \\ -10 \\ -15 \\ -20 \end{pmatrix} \in \mathbb{R}^4$$

$$\underset{\sim}{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

★

Note that $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^4$ generalise naturally to the n dimensional real space $\mathbb{R}^n$.

b) $V = P_2(\mathbb{R})$. The vector space of all real polynomials of degree at most 2.

$$\underset{\sim}{v_1} = 3 + 4x + 11x^2 \qquad \underset{\sim}{v_2} = 5 - 2x + 3x^2$$

$$\underset{\sim}{v_1} + \underset{\sim}{v_2} = 8 - 2x + 14x^2$$

$$-5\underset{\sim}{v_1} = -15 - 20x - 55x^2$$

$$\underset{\sim}{0} = 0 + 0x + 0x^2$$

★

The vector space of all real polynomials of degree at most n is denoted by $P_n(\mathbb{R})$.

c) $V = M_{23}(\mathbb{R})$. This is the vector space of all 2×3 matrices with real entries.

$$\underset{\sim}{v_1} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \end{pmatrix}, \quad \underset{\sim}{v_2} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\underset{\sim}{v_1}, \underset{\sim}{v_2} \in M_{23}(\mathbb{R})$$

$$\underset{\sim}{v_1} + \underset{\sim}{v_2} = \begin{pmatrix} 2 & 1 & 2 \\ 2 & 4 & 4 \end{pmatrix}$$

$$-5\underset{\sim}{v_1} = \begin{pmatrix} -5 & -5 & -5 \\ -10 & -15 & -20 \end{pmatrix} \in M_{23}(\mathbb{R})$$

$$\underset{\sim}{0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in M_{23}(\mathbb{R})$$

★

Note that $V = M_{mn}(\mathbb{R})$ is the vector space of all $m \times n$ matrices with real entries.

d) $V = \mathbb{C}^2$, the real vector space of all complex ordered pairs.

$$\underset{\sim}{v_1} = \begin{pmatrix} 1+i \\ 1-i \end{pmatrix} \qquad \underset{\sim}{v_2} = \begin{pmatrix} 3+4i \\ 7 \end{pmatrix}$$

$$\underset{\sim}{v_1} + \underset{\sim}{v_2} = \begin{pmatrix} 4+5i \\ 8-i \end{pmatrix} \in \mathbb{C}^2$$

$$-5\underset{\sim}{v_1} = \begin{pmatrix} -5-5i \\ -5+5i \end{pmatrix}$$

$$\underset{\sim}{0} = \begin{pmatrix} 0+0i \\ 0+0i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

#

★

$\mathbb{C}^2$ of course generalises naturally to $\mathbb{C}^n$ the real vector space of all complex n-tuples.

So a vector space is simply a collection of similar mathematical objects for which vector addition and scalar multiplication works. But what exactly do we mean by *works*? To pin this down, vector spaces are defined axiomatically. That is, we specify a number of conditions which must be met. These conditions are called axioms. Usually there are 10 vector space axioms, though some sources may bundle a few of them together.

Definition: A vector space V over the real numbers $\mathbb{R}$ is a non-empty set of objects called vectors, together with definitions of vector addition and scalar multiplication obeying the following 10 axioms:

!

1. **Closure under addition.** If $\mathbf{u}, \mathbf{v} \in V$ then $\mathbf{u} + \mathbf{v} \in V$.

2. **Associative law of addition**. If $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ then $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

3. **Commutative law of addition**. If $\mathbf{u}, \mathbf{v} \in V$ then $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

!

4. **Existence of zero**. There is a special element $\mathbf{0}$ in V called the **zero vector** which has the property that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.

5. **Existence of Negative**. For each $\mathbf{v} \in V$ there exists an element $\mathbf{w} \in V$ (the negative of $\mathbf{v}$, i.e. $\mathbf{w} = -\mathbf{v}$) such that $\mathbf{v} + \mathbf{w} = \mathbf{0}$.

!

6. **Closure under scalar multiplication**. If $\mathbf{v} \in V$ and λ is any real number (i.e. a scalar) then $\lambda\mathbf{v} \in V$.

7. **Associative law of multiplication by a scalar**. If λ, μ are any real numbers and $\mathbf{v} \in V$ then $\lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}$.

8. If $\mathbf{v} \in V$ then $1\mathbf{v} = \mathbf{v}$.

9. **Scalar distributive law**. If λ, μ are real numbers and $\mathbf{v} \in V$ then $(\lambda+\mu)\mathbf{v} = \lambda\mathbf{v}+\mu\mathbf{v}$.

10. **Vector distributive law**. If λ is a scalar and $\mathbf{u}, \mathbf{v} \in V$ then $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$.

Please note that most of the axioms are technical restrictions that just try to make sure that everything works. For example Axiom 3 says that it doesn't matter which order you add vectors, while Axiom 5 guarantees that each vector has a negative twin. Three of these Axioms are really crucial to us and we have seen two of them before:

- Closure under vector addition (Axiom 1).

- Closure under scalar multiplication (Axiom 6).

- Existence of a zero vector (Axiom 4).

Axiom 4 says something which is both subtle and obvious! Every vector space must have a zero vector. You do not need to memorise these axioms, but some of the examples in your problem sets will ask you to play around with them.

Homework: Prove that $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in V$.

It is hard to believe that this requires proof. This statement is saying that if you multiply any vector by the zero real number you will get the zero vector. Observe that the second zero is bold (vector) but the first zero is not bold (scalar). The existence of the zero vector is guaranteed by Axiom 4. There is nothing in the axioms about what happens when you multiply a vector by the zero **number** so we need to construct a proof. This is surprisingly tricky.

Proof:

Observe that $(0 + 1)\mathbf{v} = 1\mathbf{v} = \mathbf{v}$. (Axiom 8)

But it is also true that $(0 + 1)\mathbf{v} = 0\mathbf{v} + 1\mathbf{v} = 0\mathbf{v} + \mathbf{v}$. (Axiom 9)

Thus equating the right hand sides we have

$$0\mathbf{v} + \mathbf{v} = \mathbf{v}$$

At this stage you may imagine that we could cancel $\mathbf{v}$ from both sides of the equation. But cancellation doesn't exist yet, and neither does a definition of vector subtraction. But we DO know that $-\mathbf{v}$ exists. (Axiom 5). Thus adding $-\mathbf{v}$ to the right of both sides

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

yielding via Axiom 5 again

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

Axiom 2 allows us to shuffle the brackets

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

and thus

$$0\mathbf{v} + \mathbf{0} = \mathbf{0}$$

Finally applying Axiom 4 we have

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

as required.

$\bigstar$

Fortunately we almost never crawl around at the axiomatic level! We have bigger fish to fry.

Note that strictly speaking, a vector space is made up of:

1) A set of objects which we call vectors.

2) A set of scalars, which is almost always $\mathbb{R}$ but could be $\mathbb{C}$ or indeed any field. Don't worry to much about what a field is.

3) A definition of vector addition and a definition of scalar multiplication.

Generally 2) and 3) are obvious and natural, so we often describe a vector space simply as a collection of vectors and leave the other details to take care of themselves.

In this course you will not need to prove that a famous space like $V = P_2(\mathbb{R})$ is a vector space. There are 10 axioms to grind through, and $P_2(\mathbb{R})$ is already well understood. Of much more interest is whether or not particular subsets S are vector spaces.

We will close this lecture by examining how to prove that a subset S is **not** a vector space. We do this by showing that an axiom fails. Start with Axiom 4 (existence of a zero) and if necessary move on to the closure axioms 1 and 6. It is best not to tangle with the other axioms.

Proving that a set S actually **is** a vector space is a much more difficult task and will be dealt with in Lecture 3.

Example 2: Let S be the subset of $\mathbb{R}^2$ given by

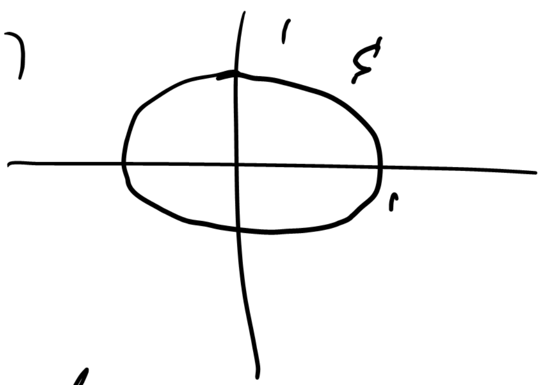
$$S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; x^2 + y^2 = 1 \right\}$$

with the usual definitions of vector addition and scalar multiplication.

a) Describe the subset S of the vector space $\mathbb{R}^2$ geometrically.

b) Prove that S is not a vector space by displaying a failure of one of the 10 vector space axioms.

a)



Circle centre $(0,0)$ radius 1

b) Consider $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$0^2 + 0^2 \neq 1$

$\therefore \begin{pmatrix} 0 \\ 0 \end{pmatrix} \notin S$

$\therefore S$ is <u>not</u> a vector space. ★

Example 3: Let S be the subset of $\mathbb{R}^2$ given by

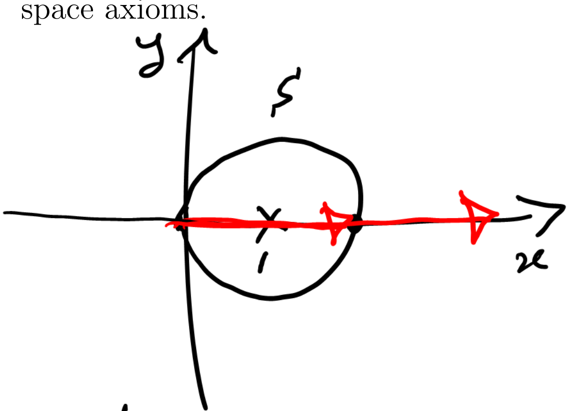
$$S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; (x-1)^2 + y^2 = 1 \right\}$$

with the usual definitions of vector addition and scalar multiplication.

a) Describe the subset S of the vector space $\mathbb{R}^2$ geometrically.

b) Prove that S is not a vector space by displaying a failure of one of the 10 vector space axioms.

a)



circle centre $(1,0)$ radius $= 1$

b) Let $\underset{\sim}{v} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$.

Then $\underset{\sim}{v} \in S$ since $(2-1)^2 + 0^2 = 1$ ✓✓

Now $3\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$

and $(6-1)^2 + 0^2 \neq 1$

$\therefore$ not closed under scalar $\times$ ★

Example 4: Let S be the subset of $P_2(\mathbb{R})$ given by

$$S = \{a + bx + cx^2 \in P_2(\mathbb{R}) \mid c = ab\}$$

with the usual definitions of vector addition and scalar multiplication.

a) Write down two non-trivial vectors in S.

b) Prove that S contains a the zero vector. (Axiom 4)

c) Prove that S is not a vector space by displaying a failure of one of the other 9 vector space axioms..

a) $\underset{\sim}{v_1} = 3 + 4x + 12x^2 \in S$

$\underset{\sim}{v_2} = 5 + 6x + 30x^2 \in S$

b) $\underset{\sim}{0} = 0 + 0x + 0x^2 \in S$ since $0 = 0 \times 0$

c) $\underset{\sim}{v_1} + \underset{\sim}{v_2} = 8 + 10x + 42x^2 \notin S$.

since $42 \neq 8 \times 10$!!

$\therefore$ Not closed under addition

$\therefore$ Not a vector space.

★

Example 5: Let S be the subset of $M_{22}(\mathbb{R})$ given by

$$S = \left\{ \begin{pmatrix} a & 0 \\ 0 & a^2 \end{pmatrix} \in M_{22}(\mathbb{R}) \right\}$$

with the usual definitions of vector addition and scalar multiplication.

a) Write down a non-trivial vector in S.

b) By considering scalar multiplication prove that S is not a vector space.

a) $\begin{pmatrix} 3 & 0 \\ 0 & 9 \end{pmatrix} \in S$

b) $-4\begin{pmatrix} 3 & 0 \\ 0 & 9 \end{pmatrix} = \begin{pmatrix} -12 & 0 \\ 0 & -36 \end{pmatrix} \notin S$

$-36 \neq (-12)^2$

Not closed under scalar $\times$

$\therefore$ Not a vector space.

★

Discussion: Let S be the set of all polynomials of degree 2. Is S a vector space? (NO)

$\underset{\sim}{v_1} = 3 + 4x + x^2 \in S$

$\underset{\sim}{v_2} = 5 + 6x - x^2 \in S$

$\underset{\sim}{v_1} + \underset{\sim}{v_2} = 8 + 10x \notin S$

$\therefore$ S is not closed under $+$.

★

Discussion: Can a vector space be a mixture of polynomials and matrices?

NO !!

★

So we prove that a subset S is **not** a vector space by showing that one of the axiomatic wheels has fallen off. In the next lecture we will look at the much more difficult question of how to prove that a set S **is** a vector space.

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 3

SUBSPACES

Milan Pahor

MATH1231 ALGEBRA

SUBSPACES

A subset S of a vector space V is said to be a **subspace** if it is a vector space in its own right.

To verify that S is a subspace you only need to prove:

(I) Closure under vector addition: $\mathbf{v}_1, \mathbf{v}_2 \in S \implies \mathbf{v}_1 + \mathbf{v}_2 \in S$.

(II) Closure under scalar multiplication: $\mathbf{v}_1 \in S$ and α a scalar $\implies \alpha\mathbf{v}_1 \in S$.

(III) The existence of a zero vector $\mathbf{0} \in S$.

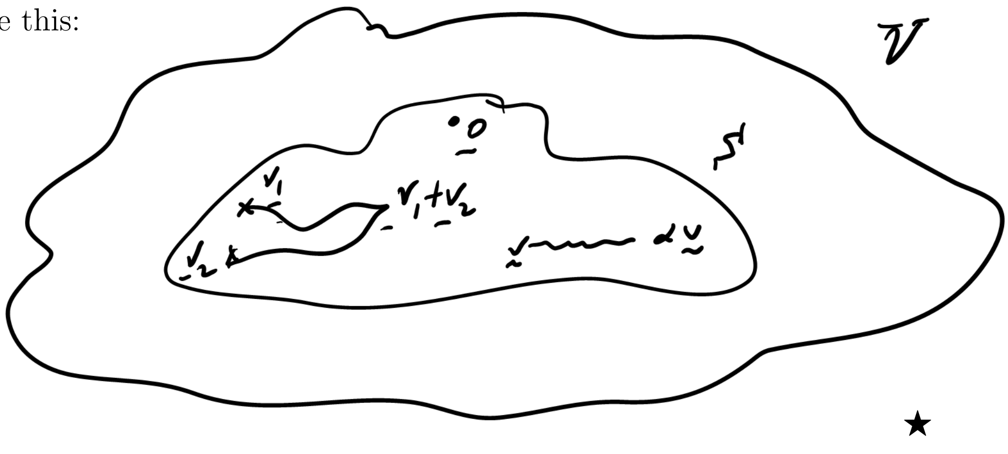
The remaining seven vector space axioms are inherited from the parent space.

You will never be asked to prove that something like $P_2(\mathbb{R})$ or $M_{34}(\mathbb{R})$ is a vector space. The proof would be a dreadful grind through the 10 vector space axioms, and these spaces are already famous for being vector spaces.

What we do ask of you however, is to verify that certain subsets of well-known vector space are also vector spaces.

Definition: A subset S of a vector space V is said to be a **subspace** if it is a vector space in its own right.

Discussion: Subspaces are little baby vector spaces living inside larger spaces. The situation looks like this:



Given a vector space V there are two very boring and obvious subspaces, $S = \{\mathbf{0}\}$ and $S = V$. That is we select either nothing or everything. These are called trivial subspaces and are of little interest.

Suppose however that S is a more interesting subset. It seems that we will still need to check all the 10 vector space axioms to prove that S is a subspace. But in fact we only need to check three!

The Subspace Theorem

A subset S of a vector space V is a **subspace** of V if and only if we have:

(I) Closure under vector addition: $\mathbf{v}_1, \mathbf{v}_2 \in S \implies \mathbf{v}_1 + \mathbf{v}_2 \in S$.

(II) Closure under scalar multiplication: $\mathbf{v}_1 \in S$ and α a scalar $\implies \alpha\mathbf{v}_1 \in S$.

(III) The existence of a zero vector $\mathbf{0} \in S$.

Proof: It is easily checked that all the other vector space properties are inherited from the parent space.

This is great news, for both students and examiners. All vector space questions can now be framed in terms of subspaces.......but they are still very hard problems!

Please look very,very carefully at the language and setup of the following proofs. You will need to phrase your arguments accurately and formally in exams.

Example 1: Let $S = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; 2x + 3y - z = 0 \right\}$.

a) Write down two vectors in S and one vector not in S.

b) Prove that S is a subspace of $\mathbb{R}^3$.

c) Describe S geometrically.

a) $\begin{pmatrix} 1 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ -18 \end{pmatrix} \in S$, $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \notin S$

b) I) Closure under addition:

Let $\underset{\sim}{v_1}, \underset{\sim}{v_2} \in S$ ← Select vectors

description
$$\left\{ \begin{array}{l} \underset{\sim}{v_1} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, \underset{\sim}{v_2} = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \\ \text{then } 2x_1 - y_1 - z_1 = 0 \text{ and } 2x_2 + 3y_2 - z_2 = 0 \end{array} \right.$$

XX $\underset{\sim}{v_1} = 2x_1 + 3y_1 - z_1 = 0$ XX

$$\underset{\sim}{v_1} + \underset{\sim}{v_2} = \begin{pmatrix} x_1+x_2 \\ y_1+y_2 \\ z_1+z_2 \end{pmatrix}$$

$$2(x_1+x_2)+3(y_1+y_2)-(z_1+z_2) \cancel{=0}$$

$$= 2x_1+2x_2+3y_1+3y_2-z_1-z_2$$

$$= (2x_1+3y_1-z_1)+(2x_2+3y_2-z_2)$$

$$= 0+0 = 0$$

$\therefore \underset{\sim}{v_1}+\underset{\sim}{v_2} \in S$ $\therefore S$ is closed under vector $+$

II) Let $\alpha \in \mathbb{R}$.

$$\alpha \underset{\sim}{v_1} = \alpha \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha y_1 \\ \alpha z_1 \end{pmatrix}$$

$$2(\alpha x_1)+3(\alpha y_1)-\alpha z_1$$

$$= \alpha(2x_1+3y_1-z_1) = \alpha(0) = 0.$$

$\therefore S$ is closed under scalar $\times$

III) Consider $\underset{\sim}{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

Then $\underset{\sim}{0} \in S$ since

$$2(0)+3(0)-0 = 0$$

$\therefore$ I, II, III $\Rightarrow S$ is a subspace of $\mathbb{R}^3$.

c) S is just a plane through the origin in $\mathbb{R}^3$.

★

Example 2: Let

$$S = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} t \;; t \in \mathbb{R} \right\}$$

That is, S is simply a particular straight line in $\mathbb{R}^3$ passing through the origin.

a) Write down two vectors in S and one vector not in S.

b) Prove that S is a subspace of $\mathbb{R}^3$.

a) $\begin{pmatrix} 6 \\ 2 \\ 10 \end{pmatrix}, \begin{pmatrix} -3 \\ -1 \\ -5 \end{pmatrix} \in S, \quad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \notin S$

b) (I) Closure under +

Let $\underset{\sim}{v_1}, \underset{\sim}{v_2} \in S$ (select vectors)

$\underset{\sim}{v_1} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} t_1, \quad \underset{\sim}{v_2} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} t_2$

where $t_1, t_2 \in \mathbb{R}$. (described)

$$\underset{\sim}{v_1} + \underset{\sim}{v_2} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} t_1 + \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} t_2 = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} (t_1 + t_2) \in S.$$

$\therefore$ S is closed under +

II) Closure under scalar $\times$

Let $\alpha \in \mathbb{R}$

$$\alpha \underset{\sim}{v_1} = \alpha \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} t_1 = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} (\alpha t_1) \in S \text{ !!}$$

$\therefore$ S is closed under scalar $\times$.

III) Show $\underset{\sim}{0} \in S$.

$$\underset{\sim}{0} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix} (0)$$

$\therefore$ $\underset{\sim}{0} \in S$.

$\therefore$ S is a subspace of $\mathbb{R}^3$

★

Observe from the previous two examples that subspaces in $\mathbb{R}^3$ (and also $\mathbb{R}^2$) are always simple straight flat thin objects! Just lines and planes through the origin. That is it!!

Remember it is easy to prove that S is NOT a subspace of a vector space. We did this at the end of the last lecture. All you need to show is something, anything going wrong. Proving that S **is** a subspace of a vector space will **always** be a battle, and will demand a general, formal proof. You cannot look at particular vectors!

Example 3: Let $S = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{22}(\mathbb{R}) \;\middle|\; a = d \right\}$.

a) Write down two vectors in S and one vector not in S. (Not part of proof!)

b) Prove that S is a subspace of $M_{22}(\mathbb{R})$.

a) $\begin{pmatrix} 3 & 7 \\ 2 & 3 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 11 & 4 \end{pmatrix} \in S \qquad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \notin S$

b) I) Closure under vector addition.

Let $\underset{\sim}{v}_1, \underset{\sim}{v}_2 \in S$ (select two vectors in S)

description $\left\{ \begin{array}{l} \underset{\sim}{v}_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}, \quad \underset{\sim}{v}_2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \\ \text{and} \quad a_1 = d_1, \quad a_2 = d_2 \end{array} \right.$

Then $\underset{\sim}{v}_1 + \underset{\sim}{v}_2 = \begin{pmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{pmatrix}$

Now $a_1 + a_2 = d_1 + d_2$

$\therefore$ S is closed under addition

II) Let $\alpha \in \mathbb{R}$, $\alpha \underset{\sim}{v}_1 = \alpha \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} = \begin{pmatrix} \alpha a_1 & \alpha b_1 \\ \alpha c_1 & \alpha d_1 \end{pmatrix}$

and $\alpha a_1 = \alpha d_1$ since $a_1 = d_1$

$\therefore$ S is closed under scalar $\times$

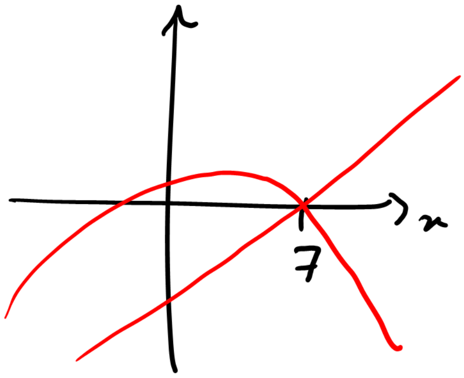
III) $\underset{\sim}{0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in S$ since $0 = 0$.

Hence S is a subspace of $M_{22}(\mathbb{R})$ ★

Example 4: Let $S = \{\mathbf{p} \in P_2(\mathbb{R}) \mid \mathbf{p}(7) = 0\}$.

a) Write down two vectors in S and one vector not in S.

b) Prove that S is a subspace of $P_2(\mathbb{R})$.



a) $(x-7)^2 = x^2 - 14x + 49$ is in S!

$(x-7) \in S$!!

$x^2+1 \notin S$

b) I) Let $p_1, p_2 \in S$

Then $p_1(7) = 0$, $p_2(7) = 0$

$$(p_1+p_2)(7) = p_1(7) + p_2(7) = 0+0 = 0.$$

$\therefore p_1 + p_2 \in S$

II) $(\alpha p_1)(7) = \alpha(p_1(7)) = \alpha(0) = 0.$

$\therefore \alpha p_1 \in S$

III) Consider the zero polynomial $\underset{\sim}{0}$

$\underset{\sim}{0} = 0x^2 + 0x + 0.$

at $x = 7$: $\underset{\sim}{0}(7) = 0$

$\therefore \underset{\sim}{0} \in S$

Hence S is a subspace of $P_2(\mathbb{R})$

★

Example 5: Let $S = \{\mathbf{p} \in P_2(\mathbb{R}) \mid \mathbf{p}(0) = 7\}$. Show that S is not a subspace of $P_2(\mathbb{R})$.

Let $p_1, p_2 \in S$

Then $p_1(0) = 7$, $p_2(0) = 7$

$(p_1+p_2)(0) = p_1(0) + p_2(0) = 7+7 = 14 \neq 7.$

$\therefore p_1 + p_2 \notin S$. So not closed under +

$\therefore$ Not a subspace.

★

OR $\underset{\sim}{0} \notin S$. $\therefore$ Not a subspace of $P_2(\mathbb{R})$

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 4

LINEAR COMBINATIONS AND SPANS IN $\mathbb{R}^n$

Milan Pahor

MATH1231 ALGEBRA

LINEAR COMBINATIONS AND SPANS IN $\mathbb{R}^n$

The span of a set of vectors $W = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ in a vector space V is the set of all possible linear combinations $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n$.

If the span of W is the whole space V we say that W spans V, or equivalently that W is a spanning set for V.

W is a spanning set for $\mathbb{R}^n \iff$ no zero rows in the associated echelon form.

Gaussian ELIMINATION

Any augmented matrix may be reduced to echelon form via the elementary row operations

$$R_i = R_i \pm \alpha R_j \qquad \text{and} \qquad R_i \leftrightarrow R_j$$

Once in echelon form the system may be solved via back-substitution.

An inconsistent equation at any stage of reduction indicates that there is no solution and you may stop. Else

If every column on the LHS is a leading column then the solution is unique. Else

The presence of a non-leading column on the LHS of the echelon form indicates infinite solutions with the non-leading variables then serving as parameters.

In the next four lectures we will turn to three crucial vector space concepts, **"span"** in this lecture, **" linear independence"**' in the next and finally the the idea of a vector space **"basis"**.

Please note that the theory of systems of linear equations and row reduction from Math1131/1141 will be the backbone of our analysis over all four classes. I will revise this theory as we go, however if you are a little rusty, you should go back and look at your previous notes. We will need to be fully aware of the various conditions which guarantee unique, infinite or no solutions for various systems of equations.

Definition: Suppose that $\{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}\}$ is a collection of n individual vectors in a vector space V and that $\lambda_1, \lambda_2, ..., \lambda_n$ are scalars. Then

$$\lambda_1 \mathbf{v_1} + \lambda_2 \mathbf{v_2} + ... + \lambda_n \mathbf{v}_n$$

is called a linear combination of the set of vectors.

Example 1: For the set of vectors $\left\{ \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} \right\} \in \mathbb{R}^3$, evaluate the linear combination $3\begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 5\begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix}$.

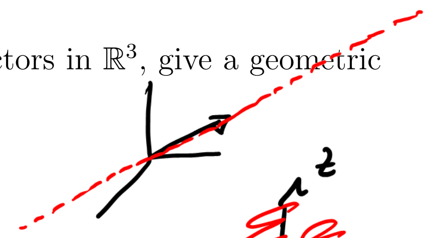
$$= \begin{pmatrix} 6 \\ 2 \\ 15 \end{pmatrix} + \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -30 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \\ -13 \end{pmatrix}$$

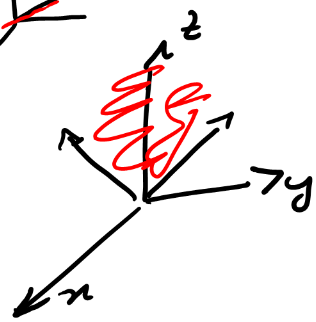
★

Definition: The **span** of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ in a vector space V, denoted by $\text{span}(\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\})$, is the set of all possible linear combinations of the given vectors.

The span of a batch of vectors in a vector space V is essentially everything that can be constructed, using those vectors as building blocks.

Example 2: For each of the following collections W of vectors in $\mathbb{R}^3$, give a geometric description of $\text{span}(W)$.





a) $W = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$. span = line through $\underset{\sim}{0}$.

b) $W = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix} \right\}$. = plane through 0

c) $W = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix} \right\}$. = line !!

!!! $\begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}$!!!

d) $W = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 6 \end{pmatrix} \right\}$. Plane : spanned by $\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix} \right\}$

e) $W = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$. all of $\mathbb{R}^3$

$\underset{\sim}{i}$ $\quad \underset{\sim}{j}$ $\quad \underset{\sim}{k}$

★

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Fact: The span of a subset of vectors in a vector space V is always a subspace of V.

So, the span of a batch of vectors W in a vector space V is always *something nice*, but we often wonder......is it the whole space V, is it *everything*? Can every vector in V be constructed from the building block vectors in W?

If the span of a set of vectors $W = \{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_n\}$ is **all** of the vector space V we say that W spans V and there is then sufficient information in W to construct the entire space V.

In the above we have only e) spanning all of $\mathbb{R}^3$. The sets in b) and d) span a plane in $\mathbb{R}^3$ while a) and c) just span a line in $\mathbb{R}^3$.

★

Example 3: Show that the vector $\begin{pmatrix} 9 \\ 8 \\ 5 \\ 18 \end{pmatrix} \in \mathbb{R}^4$ is in span of $\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 1 \\ 3 \end{pmatrix} \right\}$.

We need to find scalars α and β such that

$$\alpha \begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix} + \beta \begin{pmatrix} 3 \\ 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 9 \\ 8 \\ 5 \\ 18 \end{pmatrix}. \text{ So}$$

$$\begin{aligned} \alpha + 3\beta &= 9 \\ 2\alpha + \beta &= 8 \\ \alpha + \beta &= 5 \\ 4\alpha + 3\beta &= 18 \end{aligned}$$

which leads us to the augmented matrix

4 equations in 2 unknowns

$$\left(\begin{array}{cc|c} 1 & 3 & 9 \\ 2 & 1 & 8 \\ 1 & 1 & 5 \\ 4 & 3 & 18 \end{array} \right)$$

Observe very,very carefully how the vectors being used to construct the span are on the left of the augmented matrix, while the vector being constructed is on the right hand side. It is OK to go straight to the augmented matrix, but you must always remember to place the vectors into the augmented matrix **vertically**. You should also signal this process to the reader by announcing that you are:

"Moving to the canonical coefficient matrix".

We now solve the system by row reducing to echelon form and back-substituting as per the Gaussian elimination section of Math1131/1141.

A matrix is in echelon form is

- Each successive row has more zeros injected from the left than the row above; and
- Complete rows of zeros sit at the bottom.

When a matrix is in echelon form the first non-zero entry in each row is called a leading (or pivot) element and the columns containing leading elements are called leading (or pivot) columns.

Any augmented matrix may be reduced to echelon form via the elementary row operations

$$R_i = R_i \pm \alpha R_j \qquad \text{and} \qquad R_i \leftrightarrow R_j$$

Once in echelon form the system may be solved via back-substitution.

An inconsistent equation at any stage of reduction indicates that there is no solution and you may stop. Else

If every column on the LHS is a leading column then the solution is unique. Else

The presence of a non-leading column on the LHS of the echelon form indicates infinite solutions with the non-leading variables then serving as parameters.

So we have:

$$\left(\begin{array}{cc|c} 1 & 3 & 9 \\ 2 & 1 & 8 \\ 1 & 1 & 5 \\ 4 & 3 & 18 \end{array}\right) \xRightarrow[\substack{R_3 = R_3 - R_1 \\ R_4 = R_4 - 2R_1}]{R_2 = R_2 - 2R_1} \left(\begin{array}{cc|c} 1 & 3 & 9 \\ 0 & -5 & -10 \\ 0 & -2 & -4 \\ 0 & -9 & -18 \end{array}\right)$$

$$R_2 = -\tfrac{1}{5} R_2 \quad \left(\begin{array}{cc|c} 1 & 3 & 9 \\ 0 & 1 & 2 \\ 0 & -2 & -4 \\ 0 & -9 & -18 \end{array}\right) \xRightarrow[R_4 = R_4 + 9R_2]{R_3 = R_3 + 2R_2} \left(\begin{array}{cc|c} 1 & 3 & 9 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Every column on the LHS leads.

$\therefore$ Sol^n is unique!!

$$\beta = 2$$
$$\alpha + 3\beta = 9$$
$$\alpha + 6 = 9$$
$$\alpha = 3$$
★

$$3\begin{pmatrix} 1 \\ 2 \\ 1 \\ 4 \end{pmatrix} + 2\begin{pmatrix} 3 \\ 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 9 \\ 8 \\ 5 \\ 18 \end{pmatrix}$$

Example 4: Consider the subset $W = \left\{ \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 7 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 6 \end{pmatrix} \right\}$ of $\mathbb{R}^3$.

a) Find a condition on $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ which guarantees that $\mathbf{b}$ is in span(W).

b) Describe span(W) geometrically.

c) Does W span all of $\mathbb{R}^3$.

a) $\alpha_1 \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 3 \\ 1 \\ 7 \end{pmatrix} + \alpha_4 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \alpha_5 \begin{pmatrix} 2 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$

$$\left(\begin{array}{ccccc|c} 1 & 0 & 3 & 0 & 2 & b_1 \\ 2 & 0 & 1 & 1 & 2 & b_2 \\ 4 & 0 & 7 & 1 & 6 & b_3 \end{array}\right) \xrightarrow[R_3 = R_3 - 4R_1]{R_2 = R_2 - 2R_1}$$

$$\left(\begin{array}{ccccc|c} 1 & 0 & 3 & 0 & 2 & b_1 \\ 0 & 0 & -5 & 1 & -2 & b_2 - 2b_1 \\ 0 & 0 & -5 & 1 & -2 & b_3 - 4b_1 \end{array}\right) \xrightarrow{R_3 = R_3 - R_2}$$

$$\begin{array}{ccccc} \checkmark & \times & \checkmark & \times & \times \end{array}$$
$$\left(\begin{array}{ccccc|c} 1 & 0 & 3 & 0 & 2 & b_1 \\ 0 & 0 & -5 & 1 & -2 & b_2 - 2b_1 \\ 0 & 0 & 0 & 0 & 0 & b_3 - 2b_1 - b_2 \end{array}\right)$$

$(b_3 - 4b_1) - (b_2 - 2b_1)$

$b_3 - 4b_1 - b_2 + 2b_1$

$b_3 - 2b_1 - b_2$

Clearly $b_3 - 2b_1 - b_2 = 0$

ie $\boxed{b_3 = 2b_1 + b_2}$

b) $z = 2x + y \Rightarrow 2x + y - z = 0$

plane through origin

c) No There is a zero row in the echelon form.

★

You will observe from the above example that, as soon as there is a zero row in the echelon form, the original set of vectors will not span $\mathbb{R}^n$, since there is then a possibility of a paradox on the RHS.

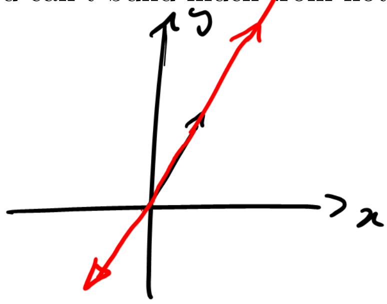
Fact: W is a spanning set for $\mathbb{R}^n \iff$ no zero rows in the associated echelon form.

Example 5: Prove that the following set of vectors does not span $\mathbb{R}^2$.

$$W = \left\{ \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 \\ 12 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -4 \end{pmatrix} \right\}$$

There is no shortage of vectors here but they are all pretty much useless! We have lots of redundant vector multiples, and the zero vector is of no value at all when it comes to spanning a space. You can't build much from nothing!

Sketch



The issue of spanning is always formally resolved via Gaussian Elimination.

To keep your examiner happy we say:

"Moving to the canonical coefficient matrix".

We then just plonk the vectors vertically into a matrix and reduce. Note that we do not have a RHS on the matrix as we are not building anything in particular, we are trying to build *everything*:

$$\begin{pmatrix} 1 & 2 & 3 & 0 & -1 \\ 4 & 8 & 12 & 0 & -4 \end{pmatrix} \xrightarrow{R_2 = R_2 - 4R_1} \begin{pmatrix} 1 & 2 & 3 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

there is a zero row

$\therefore$ Vectors do not span $\mathbb{R}^2$

They do span line through $\underline{0}$

In fact they span line $\underline{y = 4x}$

★

W is a spanning set for $\mathbb{R}^n \iff$ no zero rows in the associated echelon form.

Example 6: Prove that the following set of vectors do span $\mathbb{R}^3$.

$$W = \left\{ \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix} \right\}$$

Solution: Moving to the canonical coefficient matrix:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 4 \\ 2 & 6 & 0 \end{pmatrix} \xrightarrow[R_3 = R_3 - 2R_1]{R_2 = R_2 - R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 2 & -6 \end{pmatrix}$$

$$R_3 = R_3 - R_2 \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & -7 \end{pmatrix}$$

No zero rows in echelon form

$\therefore$ Vectors span $\mathbb{R}^3$

Make sure you mention what is significant about the echelon form.

"There are no zero rows in the echelon form, thus the vectors span $\mathbb{R}^3$ ".

★

Example 7: Suppose that two random vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in $\mathbb{R}^3$ are added to W. Will these 5 vectors still span $\mathbb{R}^3$?

There are two ways of thinking about this.

1) I still have enough information....in fact even more! So YES the larger set will still span $\mathbb{R}^3$.

2) I will still have no zero row in the echelon form as the orginal three vectors are still present! So YES the larger set will still span $\mathbb{R}^3$.

★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 5

LINEAR INDEPENDENCE

Milan Pahor

MATH1231 ALGEBRA

LINEAR INDEPENDENCE

The span of a set of vectors $W = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ in a vector space V is the set of all possible linear combinations $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n$.

If the span of W is the whole space V we say that W spans V or equivalently that W is a spanning set for V.

W is a spanning set for $\mathbb{R}^n \iff$ no zero rows in the associated echelon form.

A set of n vectors $\{\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_n\}$ in a vector space V is said to be linearly independent if

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0.$$

A set of vectors is linear independent $\iff$ every column on the LHS of the associated echelon form is a leading column.

If a non-leading column exists, the vectors are linearly dependent (that is, NOT linearly independent) and back substitution will generate a specific non-trivial linear dependence.

Non-leading vectors are trash, leading vectors are gold.

In the previous lecture we analysed the concept of **"span"** in $\mathbb{R}^n$. Before dealing with the theory of linear independence, let's look at the concept of span, as it relates to abstract vector spaces like $P_2(\mathbb{R})$ and $M_{mn}(\mathbb{R})$. It turns out that these abstract vector spaces can be dealt with using exactly the same tools, together with one simple little trick.

Example 1: Let W be the pair of vectors $W = \{1 + 2x + 4x^2, 1 - 7x + 3x^2\}$ in $P_2(\mathbb{R})$.

Prove that the vector $\mathbf{v} = 3 + 24x + 14x^2$ is in $\text{span}(W)$ and express $\mathbf{v}$ as a linear combination of the vectors in W.

Solution: Let's first do it the long way by building $3 + 24x + 14x^2$ using the two vectors in W. Assume that

$$3+24x+14x^2 = \alpha(1+2x+4x^2)+\beta(1-7x+3x^2) = \alpha+2\alpha x+4\alpha x^2+\beta-7\beta x+3\beta x^2$$

$$= (\alpha + \beta) + (2\alpha - 7\beta)x + (4\alpha + 3\beta)x^2.$$

Equating coefficients of of powers of x yields the system of equations:

$$\begin{aligned} \alpha + \beta &= 3 \\ 2\alpha - 7\beta &= 24 \\ 4\alpha + 3\beta &= 14 \end{aligned}$$

which leads us to the augmented matrix

$$\left(\begin{array}{rr|r} 1 & 1 & 3 \\ 2 & -7 & 24 \\ 4 & 3 & 14 \end{array}\right)$$

!!! BUT !!!

There is a much quicker way to get to this augmented matrix!

We can treat the quadratic $3+24x+14x^2 \in P_2(\mathbb{R})$ as if it were the vector $\begin{pmatrix} 3 \\ 24 \\ 14 \end{pmatrix} \in \mathbb{R}^3$. That is $P_2(\mathbb{R})$ is just $\mathbb{R}^3$ in disguise. We will be able to prove this formally later!

We say "identifying $a + bx + cx^2$ with the coordinate vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and moving to the canonical coefficient matrix we obtain"

$$\left(\begin{array}{rr|r} 1 & 1 & 3 \\ 2 & -7 & 24 \\ 4 & 3 & 14 \end{array}\right)$$

The shortcut is a great tool in online quizzes where working is not marked. In a formal written exam it is best to show all working, but you know exactly where you will end up!

Discussion: What if we identified $3 + 24x + 14x^2$ with $\begin{pmatrix} 14 \\ 24 \\ 3 \end{pmatrix}$ instead?

$$\left(\begin{array}{cc|c} 1 & 1 & 3 \\ 2 & -7 & 24 \\ 4 & 3 & 14 \end{array}\right) \xrightarrow[R_3 = R_3 - 4R_1]{R_2 = R_2 - 2R_1} \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & -9 & 18 \\ 0 & -1 & 2 \end{array}\right)$$

$$R_3 = R_3 - \tfrac{1}{9}R_2 \qquad \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & -9 & 18 \\ 0 & 0 & 0 \end{array}\right)$$

There is a sol^n

$\therefore \underset{\sim}{v} \in \operatorname{span}(W)$

$$-9\beta = 18 \Rightarrow \underline{\beta = -2}$$

$$\alpha + \beta = 3 \rightarrow \alpha + -2 = 3$$

$$\rightarrow \underline{\alpha = 5}$$

$$3 + 24x + 14x^2 = 5(1 + 2x + 4x^2) - 2(1 - 7x + 3x^2)$$

★ $3 + 24x + 14x^2 = 5(1 + 2x + 4x^2) - 2(1 - 7x + 3x^2)$ ★

Example 2: Does W from the previous example span $P_2(\mathbb{R})$?

NO!!

Since zero row on LHS of the echelon form!

★

Example 3: Let $W = \left\{ \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ 3 & 0 \end{pmatrix}, \begin{pmatrix} 3 & 5 \\ 6 & 1 \end{pmatrix}, \begin{pmatrix} 4 & 3 \\ 5 & 2 \end{pmatrix} \right\}$ be four vectors in $M_{22}(\mathbb{R})$.

Prove that W does not span $M_{22}(\mathbb{R})$.

Identifying $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with the coordinate vector $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$ and moving to the canonical coefficient matrix we obtain:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 3 \\ 1 & 3 & 6 & 5 \\ 0 & 0 & 1 & 2 \end{pmatrix} \xrightarrow[R_3 = R_3 - R_1]{R_2 = R_2 - R_1} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & -1 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

$$\xrightarrow{R_3 = R_3 - R_2} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix} \xrightarrow{R_4 = R_4 - R_3} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The four given matrices do <u>not</u> span $M_{22}(\mathbb{R})$ because there is a zero row in <u>echelon</u> form.

★

A more formal construction of the coefficient matrix above on the next page.

A more formal attack on Example 3

Suppose we attempt to write a random matrix in $M_{22}(\mathbb{R})$ as a linear combination of the given matrices.

$$\alpha\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + \beta\begin{pmatrix} 2 & 3 \\ 3 & 0 \end{pmatrix} + \gamma\begin{pmatrix} 3 & 5 \\ 6 & 1 \end{pmatrix} + \delta\begin{pmatrix} 4 & 3 \\ 5 & 2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then

$$\begin{pmatrix} \alpha+2\beta+3\gamma+4\delta & \alpha+3\beta+5\gamma+3\delta \\ \alpha+3\beta+6\gamma+5\delta & \gamma+2\delta \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad \text{leading to the system}$$

$$\begin{aligned} \alpha+2\beta+3\gamma+4\delta &= a \\ \alpha+3\beta+5\gamma+3\delta &= b \\ \alpha+3\beta+6\gamma+5\delta &= c \\ \gamma+2\delta &= d \end{aligned}$$

and hence the augmented matrix

$$\left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & a \\ 1 & 3 & 5 & 3 & b \\ 1 & 3 & 6 & 5 & c \\ 0 & 0 & 1 & 2 & d \end{array}\right)$$

We are not trying to construct anything in particular so we can dump the RHS and consider

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 3 \\ 1 & 3 & 6 & 5 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

This formal approach took much longer! In a written exam you should work through all these gory details. In an on-line quiz of course, you can take shortcuts.

LINEAR INDEPENDENCE

Suppose that $W = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a set of three vectors in a vector space with the property that

$$6\mathbf{v}_1 + 8\mathbf{v}_2 - 2\mathbf{v}_3 = 0.$$

$\Rightarrow 2\underset{\sim}{v}_3 = 6\underset{\sim}{v}_1 + 8\underset{\sim}{v}_2$

We then say that W is a linearly **dependent** set of vectors, since the vectors are connected in some way. In fact we can even say that

$$\mathbf{v}_3 = 3\mathbf{v}_1 + 4\mathbf{v}_2$$

that is, we can write any of the vectors in terms of the others.

If no such relation exists between vectors, we say instead that the vectors are linearly independent. Linear independence is the opposite of linear dependence.

Sometimes we just say "independent" instead of "linearly independent".

We saw earlier that a set of vectors spans the entire space if and only if there are no zero rows in the associated echelon form. We have a similar clear test for independence, except that we check on **leading columns** rather than zero rows.

Definition: A set of n vectors $\{\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_n\}$ in a vector space V is said to be linearly independent if

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

In other words a set of n vectors $\{\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_n\}$ in a vector space V is said to be linearly independent if, when you go looking for an interesting linear relationship between the vectors, you end up disappointed, and all you can say is that

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

which is of course of no interest at all since it is always true.

Example 4: Show that $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix}, \begin{pmatrix} 1 \\ 9 \\ 8 \end{pmatrix} \right\}$ is a linearly independent set in $\mathbb{R}^3$.

$$\alpha_1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 9 \\ 8 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 0$$
$$2\alpha_1 + 2\alpha_2 + 9\alpha_3 = 0$$
$$3\alpha_1 + 7\alpha_2 + 8\alpha_3 = 0$$

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 2 & 9 & 0 \\ 3 & 7 & 8 & 0 \end{array}\right) \xRightarrow[R_3 = R_3 - 3R_1]{R_2 = R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 4 & 5 & 0 \end{array}\right)$$

$$R_2 \Leftrightarrow R_3 \quad \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 4 & 5 & 0 \\ 0 & 0 & 7 & 0 \end{array}\right)$$

$$7\alpha_3 = 0 \Rightarrow \alpha_3 = 0$$
$$4\alpha_2 + 5\alpha_3 = 0 \Rightarrow 4\alpha_2 = 0$$
$$\Rightarrow \alpha_2 = 0$$
$$\alpha_1 + \alpha_2 + \alpha_3 = 0 \Rightarrow \alpha_1 + 0 + 0 = 0$$
$$\Rightarrow \alpha_1 = 0$$

$\therefore \alpha_1 = \alpha_2 = \alpha_3 = 0$

$\therefore$ Vectors are linearly independent.

Once again there is a clear shortcut to the augmented matrix, which can be used in quizzes but should not be used in written tests.

In order to only have trivial solutions, we must have no parameters! That is we must have no non-leading columns on the left hand side of the associated echelon form.

★

Fact: A set of vectors is linear **independent** $\iff$ every column on the LHS of the associated echelon form is a leading column.

Fact: If a non-leading column exists on the LHS, the vectors are **dependent** and back substitution will generate a specific non-trivial linear dependence.

Example 5: Consider the set of four vectors

$$W = \left\{ \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \\ 6 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \\ 4 \end{pmatrix} \right\} \text{ in } \mathbb{R}^3.$$

$$\alpha_1 \underset{\sim}{v_1} + \alpha_2 \underset{\sim}{v_2} + \alpha_3 \underset{\sim}{v_3} + \alpha \underset{\sim}{v_4} = \underset{\sim}{0}$$

a) Prove that the vectors are dependent using Gaussian elimination.

b) Find a particular dependence between the vectors.

c) Find a subset Z of W which is linearly **independent** and has the same span as W. too much
(That is, toss out the garbage!)

a) $\left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 4 & 8 & 1 & 5 & 0 \\ 3 & 6 & 1 & 4 & 0 \end{array}\right)$

$R_2 = R_2 - 4R_1$, $R_3 = R_3 - 3R_1$ $\Longrightarrow$ $\left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}\right)$

$$\begin{array}{c} \checkmark \quad \times \quad \checkmark \quad \times \end{array}$$

$R_3 = R_3 - R_2$ $\left(\begin{array}{cccc|c} 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$

There exist non-leading cols on LHS
$\therefore$ Vectors are linearly dependent

b) Let ~~$\alpha_2 = \mu$, $\alpha_4 = \lambda$~~

Let $\alpha_2 = 1$, $\alpha_4 = 0$

$\alpha_3 + \alpha_4 = 0 \Rightarrow \alpha_3 = 0$

$\alpha_1 + 2\alpha_2 = 0 \Rightarrow$

$\alpha_1 = -2\alpha_2 = -2(1) = -2.$

$$\therefore -2\begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix} + 1\begin{pmatrix} 2 \\ 8 \\ 6 \end{pmatrix} + 0\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 0\begin{pmatrix} 1 \\ 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

c) $Z = \left\{ \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$

$\text{span}(Z) = \text{Span}(W).$

But Z is a independent set!

$\alpha_2 = \alpha_4 = 1$

$$\star \quad -3\begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 \\ 8 \\ 6 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \star$$

The leading vectors are crucial and must be kept. The non-leading vectors are trash.

Discussion: Can you see a really easy way of showing that W is a dependent set?

Example 6: Consider the set of vectors

$$W = \left\{ \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 \\ 12 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 6 \\ 10 \end{pmatrix}, \begin{pmatrix} -1 \\ -4 \end{pmatrix}, \begin{pmatrix} -3 \\ -5 \end{pmatrix} \right\} \text{ in } \mathbb{R}^2.$$

a) Prove that W spans $\mathbb{R}^2$.

b) Prove that W is a linearly dependent set using your echelon form from a).

c) Prove that W is a linearly dependent by inspection of W.

d) Find a subset Z of W which is linearly **independent** and has the same span as W. That is, get rid of the rubbish. If there are redundant vectors you have dependence!

Moving to the canonical coefficient matrix we have:

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 3 & 6 & -1 & -3 \\ 4 & 8 & 12 & 0 & 5 & 10 & -4 & -5 \end{pmatrix}$$

$$R_2 = R_2 - 4R_1 \quad \begin{array}{cccccccc} \checkmark & \times & \times & \times & \checkmark & \times & \times & \times \end{array}$$

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 3 & 6 & -1 & -3 \\ 0 & 0 & 0 & 0 & -7 & -14 & 0 & 7 \end{pmatrix}$$

a) Spans $\mathbb{R}^2$ since no zero row in echelon form.

b) W is a dependent set since there exist non-leading columns.

c) $\underset{\sim}{v}_3 = 3\underset{\sim}{v}_1 \quad \therefore$ Dependent set.

d) $Z = \left\{ \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \end{pmatrix} \right\}$

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 6

VECTOR SPACE BASES

Milan Pahor

MATH1231 ALGEBRA
VECTOR SPACE BASES

The span of a set of vectors $W = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ in a vector space V is the set of all possible linear combinations $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n$.

If the span of W is the whole space V we say that W spans V or equivalently that W is a spanning set for V.

W is a spanning set for $\mathbb{R}^n \iff$ no zero rows in the associated echelon form.

A set of n vectors $\{\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_n\}$ in a vector space V is said to be linearly independent if

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0.$$

A set of vectors is linear independent $\iff$ every column on the LHS of the associated echelon form is a leading column.

If a non-leading column exists, the vectors are linearly dependent (that is, NOT linearly independent) and back substitution will generate a specific non-trivial linear dependence.

Non-leading vectors are trash, leading vectors are gold.

A **basis** for a vector space is a linearly independent spanning set. So

A set of vectors is a **basis** for $\mathbb{R}^n \iff$ the associated echelon form has every column on the LHS leading **AND** no zero rows.

The **dimension** of a vector space is the number of vectors in any basis.

This lecture and the next will introduce the critical concept of **basis** and **dimension** in a vector space. But first two examples dealing with linear independence (and span) in abstract vector spaces.

Example 1: Consider the subset of three vectors $W = \{1+2x, 1-x, -1+10x\}$ in $P_1(\mathbb{R})$.

Recall that $P_1(\mathbb{R})$ is the vector space of all polynomials of degree at most 1.

a) Prove that the three vectors are linearly dependent.

b) Find a dependence.

c) Prove that the three vectors span $P_1(\mathbb{R})$.

We will do the question formally, but also point out the shortcut. The shortcut may be useful in online quizzes or if you are squeezed for time, while the formal approach should be used in written exams.

a)

$\alpha_1(1+2x) + \alpha_2(1-x) + \alpha_3(-1+10x) = \underline{0}$

$\alpha_1 + 2\alpha_1 x + \alpha_2 - \alpha_2 x - \alpha_3 + 10\alpha_3 x = \underset{\sim}{0}$

$(\alpha_1 + \alpha_2 - \alpha_3) + (2\alpha_1 - \alpha_2 + 10\alpha_3)x = 0 + 0x$

$\alpha_1 + \alpha_2 - \alpha_3 = 0$

$2\alpha_1 - \alpha_2 + 10\alpha_3 = 0$

$$\left(\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 2 & -1 & 10 & 0 \end{array}\right)$$

$R_2 = R_2 - 2R_1$

$$\begin{array}{cccc} \checkmark & \checkmark & \times & \\ \left(\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & -3 & 12 & 0 \end{array}\right) \end{array}$$

There exists a non-leading column $\therefore$ Dependent.

b) ~~Let $\alpha_3 = t$~~

Let $\underline{\alpha_3 = 1}$

$-3\alpha_2 + 12\alpha_3 = 0$

$-3\alpha_2 + 12 = 0$

$\underline{\alpha_2 = 4}$

$\alpha_1 + \alpha_2 - \alpha_3 = 0$

$\alpha_1 + 4 - 1 = 0 \Rightarrow \alpha_1 = -3$

$-3(1+2x) + 4(1-x) + (-1+10x) = 0 + 0x$

$-3\underline{v_1} + 4\underline{v_2} + \underline{v_3} = \underline{0}$

c) Vectors span $P_1(\mathbb{R})$ since there is no zero row in echelon form.

★

We will formally define dimension soon, but at this stage you can accept that $P_1(\mathbb{R})$ is a *small* vector space. Observe that if you have lots of vectors in a tiny vector space like $P_1(\mathbb{R})$, spanning is easy (there is lots of data) but independence is hard (the space becomes crowded)!

Example 2: Consider the subset of two vectors $W = \left\{ \underset{\underset{\sim}{v_1}}{\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \end{pmatrix}}, \underset{\underset{\sim}{v_2}}{\begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}} \right\}$ in the quite LARGE vector space $M_{23}(\mathbb{R})$ of all 2×3 real matrices.

a) Prove that W is a linearly independent set.

b) Prove that W does not span $M_{23}(\mathbb{R})$.

---

a) Assume that a linear combination of the two vectors is zero.

$$\alpha_1 \underset{\sim}{v_1} + \alpha_2 \underset{\sim}{v_2} = \underset{\sim}{0}$$

$$\alpha_1 \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \alpha_1+\alpha_2 & \alpha_2 & \alpha_2 \\ 3\alpha_1+2\alpha_2 & \alpha_1+\alpha_2 & \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{aligned} \alpha_1+\alpha_2 &= 0 \\ \alpha_2 &= 0 \\ \alpha_2 &= 0 \\ 3\alpha_1+2\alpha_2 &= 0 \\ \alpha_1+\alpha_2 &= 0 \\ \alpha_2 &= 0 \end{aligned}$$

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

$R_4 = R_4 - 3R_1$
$R_5 = R_5 - R_1$

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right)$$

$R_3 = R_3 - R_2$
$R_4 = R_4 + R_2$
$R_6 = R_6 - R_2$

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

every column on LHS leads $\therefore$ linearly ind. set.

b) W fails to span $M_{23}(\mathbb{R})$ because of zero rows in echelon form

Observe that if you have only a few vectors in a big vector space like $M_{23}(\mathbb{R})$, spanning is hard (there is insufficient data) but independence is easy (there is lots of room to move about)!

BASIS

Recall that in the three dimensional vector space $\mathbb{R}^3$, the set B of three individual vectors

$$B = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

is crucial. In fact, all of the infinitely many vectors in $\mathbb{R}^3$ can be built from this pathetic little subset using nothing but linear combinations. A set like B, which provides an efficient constructive framework for a vector space is called a **basis** for the space.

A basis B for a vector space V needs to do two things:

It must hold enough structure to build the entire space.......it needs to span V.

It must not contain redundant information, it needs to be light........that is linearly independent.

This leads to the definition:

Definition: A collection of vectors B is called a **basis** for a vector space V, if B is a linearly independent spanning set for V.

Recall that issues of linear independence and spanning in $\mathbb{R}^n$ are ultimately decided by properties of the associated echelon form. If every column leads we have independence. If there are no zero rows the set spans the entire space. Thus:

Fact: For a set of vectors to be a basis for $\mathbb{R}^n$ the associated echelon form must have no zero rows and no non-leading columns.

This means the echelon form needs to look neat and square.

Example 3: Prove that $B = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 4 \end{pmatrix} \right\}$ is a basis for $\mathbb{R}^3$.

We will check linear independence first and then spanning:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 4 \end{pmatrix}$$

$R_2 = R_2 - R_1$ $$\begin{pmatrix} 1 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 4 \end{pmatrix}$$

✓ ✓ ✓

$R_3 = R_3 + \frac{1}{2}R_2$ $$\begin{pmatrix} 1 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

every col. leads => linearly independent

no zero rows => spans $\mathbb{R}^3$

$\therefore$ B is a basis for $\mathbb{R}^3$

★

What we have shown in example 3 is that the three vectors

$$B = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 4 \end{pmatrix} \right\}$$

may be efficiently used to construct the entire space $\mathbb{R}^3$.

There is just the right amount of information available in the basis B.

How many bases (plural of basis) does $\mathbb{R}^3$ have? Well there are infinitely many of course! Pretty much any three vectors pointing in different directions will do the job!

BUT

Any basis for $\mathbb{R}^3$ will have exactly three elements......no more and no less!

Fact: A vector space will admit many different bases, but each basis will have exactly the same number of elements.

Definition: The **dimension** of a vector space is the number of vectors in any of its bases.

The higher the dimension the more complicated the space.

Discussion: What is the maximum dimension a vector space can have?

Most vector spaces have very obvious bases called standard bases.

Example 4: Write down the standard basis B for each of the following vector spaces V. In each case state the dimension of the space:

a) $V = \mathbb{R}^3$. $\quad B = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ $\quad \text{Dim}(\mathbb{R}^3) = 3$.

b) $V = P_3(\mathbb{R})$ $\quad B = \{1, x, x^2, x^3\}$ $\quad \text{Dim}(P_3(\mathbb{R})) = 4$. In general $\text{Dim}(P_n(\mathbb{R})) = n + 1$.

c) $V = M_{23}(\mathbb{R})$ (the set of all real 2×3 matrices).

$$B = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

$\text{Dim}(M_{23}(\mathbb{R})) = 6$.

In general $\text{Dim}(M_{mn}(\mathbb{R})) = mn$. That is, there are mn slots to fill.

★

Example 5: What is the dimension of each of the following real vector spaces V?

a) $V = P_7(\mathbb{R})$. $\quad \text{Dim}(P_7(\mathbb{R})) = 8$

b) $V = M_{35}(\mathbb{R})$. $\quad \text{Dim}(M_{35}(\mathbb{R})) = 15$

c) $\quad V = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \cancel{\begin{pmatrix} 2 \\ 0 \\ 0 \\ 2 \end{pmatrix}}, \cancel{\begin{pmatrix} 3 \\ 0 \\ 0 \\ 3 \end{pmatrix}}, \cancel{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$.

$\therefore \text{Dim}(V) = 2$

Note $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$ <u>is</u> a basis for span V

★ *a)* 8 *b)* 15 *c)* 2 ★

Recall that issues of linear independence and spanning in $\mathbb{R}^n$ are ultimately decided by properties of the associated echelon form. If every column leads we have independence. If there are no zero rows the set spans $\mathbb{R}^n$. Thus for a set of vectors to be a basis for $\mathbb{R}^n$ the associated echelon form must have no zero rows and no non-leading columns. Remember that the vectors **always** end up being placed vertically into the augmented matrix.

Example 6: Suppose that a set S of m vectors in $\mathbb{R}^n$ have been assembled vertically into a matrix and then reduced to the following echelon forms. In each case determine:

i) The number of vectors m and the dimension of the space n.

ii) Whether the m vectors are linearly independent.

iii) Whether the m vectors span $\mathbb{R}^n$.

iv) Whether the m vectors from a basis for $\mathbb{R}^n$.

✓ ✓ ✓

a) $\begin{pmatrix} 1 & 4 & 7 \\ 0 & 3 & -1 \\ 0 & 0 & 8 \end{pmatrix}$

3 vectors in $\mathbb{R}^3$
span ✓ independent ✓
∴ Basis for $\mathbb{R}^3$ ✓

✓ ✓ ✓ ✗

b) $\begin{pmatrix} 5 & 3 & 7 & 4 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 8 & 9 \end{pmatrix}$

4 vectors in $\mathbb{R}^3$
span ✓ independent ✗
Basis for $\mathbb{R}^3$ ✗

✗ ✓ ✓

c) $\begin{pmatrix} 0 & 5 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$

3 vectors in $\mathbb{R}^3$
span ✗ independent ✗
∴ Not a basis for $\mathbb{R}^3$

✓ ✗ ✓

d) $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}$

3 vectors in $\mathbb{R}^3$
span ✗ independent ✗
∴ Not a basis for $\mathbb{R}^3$

✓ ✗ ✓ ✗ ✗

e) $\begin{pmatrix} 3 & 5 & 1 & 0 & 2 \\ 0 & 0 & -1 & 8 & 1 \end{pmatrix}$

5 vectors in $\mathbb{R}^2$
span ✓ ind ✗
∴ Not a basis for $\mathbb{R}^2$

✓ ✓

f) $\begin{pmatrix} 1 & 2 \\ 0 & 4 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$

2 vectors in $\mathbb{R}^5$
spans ✗ independent ✓
∴ Not a basis for $\mathbb{R}^5$

a) $i) m = 3, n = 3$ $ii)$ Lin. Ind.=Yes $iii)$ Spans=Yes $iv)$ Basis=Yes.
b) $i) m = 4, n = 3$ $ii)$ Lin. Ind.=No $iii)$ Spans=Yes $iv)$ Basis=No.
c) $i) m = 3, n = 3$ $ii)$ Lin. Ind.=No $iii)$ Spans=No $iv)$ Basis=No.
d) $i) m = 3, n = 3$ $ii)$ Lin. Ind.=No $iii)$ Spans=No $iv)$ Basis=No.
e) $i) m = 5, n = 2$ $ii)$ Lin. Ind.=No $iii)$ Spans=Yes $iv)$ Basis=No.
f) $i) m = 2, n = 5$ $ii)$ Lin. Ind.=Yes $iii)$ Spans=No $iv)$ Basis=No.

★

Example 7: Prove that $B = \{1 - t, 1 + t, t^2 - t^3, t^2\}$ is a basis for $P_3(\mathbb{R})$.

Formal Way for Written Exams:

We start by checking linear independence:

Assume that $\alpha(1-t) + \beta(1+t) + \gamma(t^2 - t^3) + \delta t^2 = 0 + 0t + 0t^2 + 0t^3$.

Then

$\alpha - \alpha t + \beta + \beta t + \gamma t^2 - \gamma t^3 + \delta t^2 = 0 + 0t + 0t^2 + 0t^3$. Hence

$(\alpha + \beta) + (-\alpha + \beta)t + (\gamma + \delta)t^2 - \gamma t^3 = 0 + 0t + 0t^2 + 0t^3$. Leading to the system

$$\begin{array}{rrcl} \alpha + \beta & & = & 0 \\ -\alpha + \beta & & = & 0 \\ & \gamma + \delta & = & 0 \\ & -\gamma & = & 0 \end{array}$$

and hence the augmented matrix

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

Quick Way for Quizzes:

"identifying $a + bt + ct^2 + dt^3$ with the coordinate vector $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$ and moving to the canonical coefficient matrix we obtain"

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

Either way we now need to reduce......

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \xrightarrow[R_4=R_4+R_3]{R_2=R_2+R_1} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The polynomials are linearly independent as all the columns in the echelon form are leading columns. The polynomials span the entire space since there are no zero rows in the echelon form. Hence B is a basis for $P_3(\mathbb{R})$.

★

Example 8: Without using row reduction, explain why $B = \left\{ \begin{pmatrix} 1 & -2 \\ 7 & 11 \end{pmatrix}, \begin{pmatrix} 3 & -2 \\ 11 & 13 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \right\}$ is not a basis for the real vector space $M_{22}(\mathbb{R})$.

$M_{22}(\mathbb{R})$ is $\underline{4}$ dimensional

$\therefore$ B is <u>not</u> a basis, since there are only 3 vectors in B!

★

In the next lecture we will look at a range of examples dealing with linear independence, span and basis.

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 7

SPAN, INDEPENDENCE AND BASIS

Milan Pahor

MATH1231 ALGEBRA

SPAN, INDEPENDENCE AND BASIS

The span of a set of vectors $W = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ in a vector space V is the set of all possible linear combinations $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n$.

If the span of W is the whole space V we say that W spans V or equivalently that W is a spanning set for V.

W is a spanning set for $\mathbb{R}^n \iff$ no zero rows in the associated echelon form.

A set of n vectors $\{\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_n\}$ in a vector space V is said to be linearly independent if

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0.$$

A set of vectors is linear independent $\iff$ every column on the LHS of the associated echelon form is a leading column.

If a non-leading column exists, the vectors are linearly dependent (that is, NOT linearly independent) and back substitution will generate a specific non-trivial linear dependence.

Non-leading vectors are trash, leading vectors are gold.

A **basis** for a vector space is a linearly independent spanning set. So

A set of vectors is a **basis** for $\mathbb{R}^n \iff$ the associated echelon form has every column on the LHS leading **AND** no zero rows.

The **dimension** of a vector space is the number of vectors in any basis.

In this lecture we will close off the basic theory of span, independence and basis by looking at a variety of exam type questions. Note that in the next lecture, we will build upon this content and start on the theory of linear transformations between vector spaces. Almost the entire Math1231 algebra syllabus is about vector spaces in one form or another.

Recall that a basis B for an n dimensional vector space V is a linearly independent spanning set. A basis is a perfect set of vectors from which the vector space may be constructed. The spanning guarantees that there is enough information within B to construct all of V, while the independence ensures that we are not carrying around redundant rubbish in B.

The dimension of a vector space is the number of vectors in **any** basis. If V is of dimension n, any set of vectors with more than n elements will be dependent, while any set with less than n elements will fail to span the entire space. Linearly independent sets tend to be small in number, spanning sets tend to be large in number. A basis sits right in the middle. It is the Goldilocks set.

When looking for bases, our focus is on leading columns, since the non-leading vectors are trash.

Example 1: Consider the following set W of five vectors in $\mathbb{R}^3$.

$$W = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 6 \\ 8 \\ 12 \end{pmatrix} \right\}$$

a) Could a set of 5 vectors span $\mathbb{R}^3$? Yes, there is plenty to choose from.

b) Must a set of 5 vectors span $\mathbb{R}^3$? No, they could all be parallel!

c) Could a set of 5 vectors be a linear independent set in $\mathbb{R}^3$? No, too many!

d) Could a set of 5 vectors be a basis for $\mathbb{R}^3$? No, must be 3 vectors!

e) Find a basis for span(W). (Note f) will change the way we do e))

f) Find a basis for $\mathbb{R}^3$ containing as many of the vectors in W as possible.

$$\left(\begin{array}{ccccc:ccc} 1 & 2 & 1 & 2 & 6 & 1 & 0 & 0 \\ 2 & 4 & 0 & 2 & 8 & 0 & 1 & 0 \\ 3 & 6 & 0 & 3 & 12 & 0 & 0 & 1 \end{array}\right) \xrightarrow[R_3 = R_3 - 3R_1]{R_2 = R_2 - 2R_1}$$

(e) (f)

$$\left(\begin{array}{ccccc:ccc} 1 & 2 & 1 & 2 & 6 & 1 & 0 & 0 \\ 0 & 0 & -2 & -2 & -4 & -2 & 1 & 0 \\ 0 & 0 & -3 & -3 & -6 & -3 & 0 & 0 \end{array}\right) \xrightarrow{R_3 = R_3 - \frac{3}{2} R_2}$$

$$\begin{array}{cccccccc} \checkmark & \times & \checkmark & \times & \times & \times & \checkmark & \times \end{array}$$

$$\left(\begin{array}{ccccc:ccc} 1 & 2 & 1 & 2 & 6 & 1 & 0 & 0 \\ 0 & 0 & -2 & -2 & -4 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{3}{2} & 0 \end{array}\right)$$

e) Basis for span(W) $= \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$

span(W) is 2-dimensional subspace of $\mathbb{R}^3$

ie a <u>plane</u>

f) $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$

★

Example 2: The subspace of $\mathbb{R}^3$ given by

$$S = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; x + 2y + 5z = 0 \right\}$$

is a plane through the origin.

a) Find a basis B for S.

b) Expand B to a basis for $\mathbb{R}^3$.

$x + 2y + 5z = 0$

a) Let $y = \mu$, $z = \lambda$

$\therefore \quad x = -2y - 5z = -2\mu - 5\lambda$

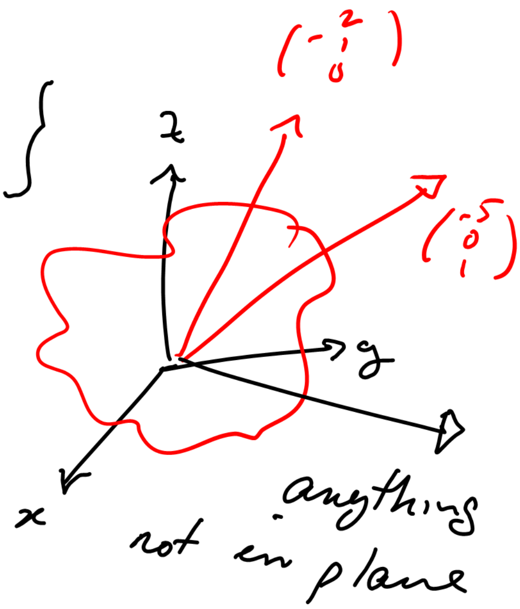
$$\therefore \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}\mu + \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix}\lambda, \quad \mu, \lambda \in \mathbb{R}$$

This parametric vector form

$\therefore$ Basis for $S = \left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix} \right\}$

b) $\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$

$\notin$ plane since
$x + 2y + 5z = 1 + 0 + 0 \neq 0$



★ $a)$ $B = \left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix} \right\}$ $b)$ Add in any vector not in S. ★

Remember that any basis for an n dimensional vector space must contain **exactly** n vectors, no more, no less. It can be further shown (see School notes) that if you have the right number of vectors, you do not need to check both spanning and independence, just one of the two will do.

Example 3: Without making any calculations prove that $B = \{1 + x,\, 1 - x,\, 3 + 4x - 5x^2,\, 10x^2\}$ is not a basis for $P_2(\mathbb{R})$.

$P_2(\mathbb{R})$ is three dimensional $\quad \{1, x, x^2\} \equiv$ Basis
B has 4 vectors.!! $\therefore$ Not a basis for $P_2(\mathbb{R})$★
too many vectors!

Example 4: Identify each of the following statements as being either true or false.

a) Any set of 5 vectors in $\mathbb{R}^4$ is linearly dependent.

True, Too many vectors!

b) Any set of 5 vectors in $\mathbb{R}^4$ spans $\mathbb{R}^4$.

False, They could be // !!

c) Some sets of 5 vectors in $\mathbb{R}^4$ span $\mathbb{R}^4$.

True, as long as they do the job!

d) Some sets of 3 vectors in $\mathbb{R}^4$ span $\mathbb{R}^4$.

False, Three are <u>not</u> enough vectors.

e) Any set of 3 vectors in $\mathbb{R}^4$ is linearly independent.

False, they <u>could</u> be //.

f) Some sets of 3 vectors in $\mathbb{R}^4$ are linearly independent.

True. But they need to point in different directions.

g) Any set of 4 linearly independent vectors in $\mathbb{R}^4$ is a basis for $\mathbb{R}^4$.

<u>True</u>. If you have the <u>right</u> <u>number</u> of vectors, you don't need <u>both</u> span and independence, either one will do! ★

Example 5: Prove that a linearly independent set cannot contain the zero vector.

Proof: Consider a set of vectors which contains the zero vector:

$$\{\mathbf{0}, \mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}\}$$

Then $17\underset{\sim}{0} + 0\underset{\sim}{v_1} + 0\underset{\sim}{v_2} + \ldots + 0\underset{\sim}{v_n} = \underset{\sim}{0}$ $*$

$\therefore$ vectors are dependent and $*$ is a dependence.

$\bigstar$

Example 6: Suppose that $B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ is a basis for a vector space V and that $\mathbf{w}$ is any vector in V. Prove that $\mathbf{w}$ can be expressed as a linear combination of the vectors in B in exactly one way.

Proof:

There exists scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that

$$\mathbf{w} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$$

Why? (B spans V).

Suppose now that it is also true that

$$\mathbf{w} = \beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \cdots + \beta_n\mathbf{v}_n$$

We have $\alpha_1\underset{\sim}{v_1} + \alpha_2\underset{\sim}{v_2} + \ldots + \alpha_n\underset{\sim}{v_n} = \beta_1\underset{\sim}{v_1} + \beta_2\underset{\sim}{v_2} + \ldots + \beta_n\underset{\sim}{v_n}$

$(\alpha_1 - \beta_1)\underset{\sim}{v_1} + (\alpha_2 - \beta_2)\underset{\sim}{v_2} + \ldots + (\alpha_n - \beta_n)\underset{\sim}{v_n} = \underset{\sim}{0}$.

$\Rightarrow \alpha_1 - \beta_1 = 0, \alpha_2 - \beta_2 = 0, \ldots (\alpha_n - \beta_n) = 0$
since B is a linearly independent set!

$\Rightarrow \alpha_1 = \beta_1, \alpha_2 = \beta_2, \ldots \alpha_n = \beta_n$!

$\bigstar$

$\therefore \underset{\sim}{w}$ can be written uniquely!

We have seen in the last few lectures that we can group vectors vertically into columns and then treat them as a single matrix. We can also go the others way, and split the columns of a matrix A up into a batch vertical vectors. The span of those vectors is then called the column space of A and denoted by $\mathrm{col}(A)$.

Example 7: Find a basis for, and the dimension of the column space of the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 1 & 2 & 0 & 3 & 0 \end{pmatrix}$$

This question is asking you to perceive the columns of A as individual vectors, and then consider a basis for the span of those vectors.....this is easy!

$$\mathrm{col}(A) = \mathrm{Sp}\left[\left\{ \begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \begin{pmatrix}2\\0\\0\\2\end{pmatrix}, \begin{pmatrix}3\\1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\4\\4\\3\end{pmatrix}, \begin{pmatrix}0\\0\\0\\0\end{pmatrix} \right\}\right]$$

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 1 & 2 & 0 & 3 & 0 \end{pmatrix} \xrightarrow{R_4 = R_4 - R_1} \begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & -3 & 3 & 0 \end{pmatrix}$$

$R_3 = R_3 - R_2$

$R_4 = R_4 + 3R_2$

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 15 & 0 \end{pmatrix}$$

$R_3 \Leftrightarrow R_4 \Rightarrow$

$$\begin{array}{ccccc} \checkmark & \times & \checkmark & \checkmark & \times \end{array}$$

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 15 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\therefore \text{Basis } \mathrm{col}(A) = \left\{ \begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \begin{pmatrix}3\\1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\4\\4\\3\end{pmatrix} \right\} \star$$

$\therefore$ Dimension $= \underline{\underline{3}}$

Example 8: Let $W = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 2 & 8 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 2 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ be a set of six vectors in $M_{22}(\mathbb{R})$.

a) Prove that the vectors are linearly dependent.

b) Prove that the vectors span $M_{22}(\mathbb{R})$.

c) Is W a basis for $M_{22}(\mathbb{R})$?

d) Find a basis for $M_{22}(\mathbb{R})$ made up of vectors from W.

---

a) This is trivial! There are simply too many vectors! The vector space $M_{22}(\mathbb{R})$ is four dimensional hence **any** set of six vectors will be dependent.

b) Spanning is harder to analyse. Maybe the vectors span, maybe they don't.

A formal approach: Let us try to build a random vector in $M_{22}(\mathbb{R})$:

$$\alpha_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix} + \alpha_4 \begin{pmatrix} 0 & 0 \\ 2 & 8 \end{pmatrix} + \alpha_5 \begin{pmatrix} 0 & 0 \\ 2 & 2 \end{pmatrix} + \alpha_6 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then

$$\begin{pmatrix} \alpha_1 + \alpha_2 + 2\alpha_3 & \alpha_2 + \alpha_3 \\ 2\alpha_4 + 2\alpha_5 & 8\alpha_4 + 2\alpha_5 + \alpha_6 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

leading to the system of equations:

$$\begin{aligned} \alpha_1 + \alpha_2 + 2\alpha_3 \qquad\qquad &= a \\ \alpha_2 + \alpha_3 \qquad\qquad &= b \\ 2\alpha_4 + 2\alpha_5 \quad &= c \\ 8\alpha_4 + 2\alpha_5 + \alpha_6 &= d \end{aligned}$$

and hence the canonical coefficient matrix:

$$\left(\begin{array}{cccccc|c} 1 & 1 & 2 & 0 & 0 & 0 & a \\ 0 & 1 & 1 & 0 & 0 & 0 & b \\ 0 & 0 & 0 & 2 & 2 & 0 & c \\ 0 & 0 & 0 & 8 & 2 & 1 & d \end{array} \right)$$

OR much more quickly:

Identifying $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$ and moving to the canonical coefficient matrix we have:

$$\left(\begin{array}{cccccc|c} 1 & 1 & 2 & 0 & 0 & 0 & a \\ 0 & 1 & 1 & 0 & 0 & 0 & b \\ 0 & 0 & 0 & 2 & 2 & 0 & c \\ 0 & 0 & 0 & 8 & 2 & 1 & d \end{array}\right)$$

$R_4 = R_4 - 4R_3$

$$\begin{array}{cccccc} \checkmark & \checkmark & \times & \checkmark & \checkmark & \times \end{array}$$
$$\begin{pmatrix} 1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 & -6 & 1 \end{pmatrix}$$

b) Vector span $M_{22}(\mathbb{R})$ since there is no zero row in echelon form

c) No, there are too many vectors.

OR No, non-leading cols $\Rightarrow$ Dependent set

d) $\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 2 & 8 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 2 & 2 \end{pmatrix} \right.$

is a basis for $M_{22}(\mathbb{R})$.

Observe in the echelon form above, that there are heaps of non-leading columns on the LHS of the augmented matrix, implying dependence of the vectors. But we already knew that. There are too many of them!

★

Example 9: Suppose that B is any set of n **orthogonal** vectors in an n dimensional vector space V. Explain briefly why B will be a basis for V.

Well we have the right number of vectors so we only need to check independence. But perpendicular vectors will certainly point in different directions! They can't depend on each other. So B is a basis for V.

Homework: Prove formally that a set of perpendicular vectors in $\mathbb{R}^n$ will always be linearly independent.

Proof:

Let $B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be a set of perpendicular vectors and suppose that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0} \qquad (*)$$

Hint: What happens when you dot both sides of (*) with $\mathbf{v}_1$?

$$(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots \alpha_n \mathbf{v}_n)\cdot(\mathbf{v}_1) = \mathbf{0}\cdot\mathbf{v}_1 = 0.$$

$$\Rightarrow \alpha_1 \underbrace{(\mathbf{v}_1\cdot\mathbf{v}_1)}_{\neq 0.} + \alpha_2 \cancel{\mathbf{v}_2\cdot\mathbf{v}_1} + \cdots \alpha_n \cancel{\mathbf{v}_n\cdot\mathbf{v}_1} = 0$$

$$\Rightarrow \alpha_1(\mathbf{v}_1\cdot\mathbf{v}_1) = 0 \Rightarrow \alpha_1 = 0.$$

Similarly $\alpha_2 = \alpha_3 = \cdots \alpha_n = 0$

$\therefore$ $\alpha_1 = \alpha_2 = \cdots \alpha_n = 0.$

★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 8

LINEAR TRANSFORMATIONS

Milan Pahor

MATH1231 ALGEBRA

LINEAR TRANSFORMATIONS

A transformation T from the vector space V into the vector space W is linear if the following two conditions are satisfied:

(I) $\quad T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$

and

(II) $\quad T(\alpha\mathbf{v}_1) = \alpha T(\mathbf{v}_1)$

for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and scalars α.

LINEAR TRANSFORMATIONS

The last seven lectures have involved a detailed analysis of the theory of vector spaces. We have looked at concrete spaces like $\mathbb{R}^3$ as well as abstract polynomial spaces like $P_2(\mathbb{R})$.

We turn now to the issue of transforming one vector space into another vetor space. This is especially important when manipulating $\mathbb{R}^3$ and $\mathbb{R}^2$. For example when you take any photo, you are transforming three dimensional space into two dimensional space.

We ensure that these transformations (also called maps) respect the underlying vector space structures by demanding **linearity**. The condition of linearity is extremely strong, guaranteeing a host of additional properties and structures. Most (but not all) of the interesting transformations between vectors spaces: reflections, rotations, projections etc are linear.

Let's start by having a look at how these transformations actually work.

Example 1: Let the transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2y \\ 3x \\ x+y \end{pmatrix}.$$

Evaluate $T\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $T\begin{pmatrix} 5 \\ 6 \end{pmatrix}$ and $T\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$T\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}$ "T maps $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} 2 \\ 3 \\ 2 \end{pmatrix}$".

$T\begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 12 \\ 15 \\ 11 \end{pmatrix}$

$T\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

★

Observe how the transformation T above, does in fact take elements of $\mathbb{R}^2$ and changes (transforms!) them into elements of $\mathbb{R}^3$. In fact, it seems to do this with ease.

If $T : V \to W$ we say that V is the domain of T and W is the codomain of T, just like with functions in Math1131. Indeed functions from high school are just transformations from the vector space $\mathbb{R}$ to the vector space $\mathbb{R}$!

It is also possible to transform abstract spaces.

Example 2: Define $T : P_2(\mathbb{R}) \to M_{22}(\mathbb{R})$ by $T(a + bx + cx^2) = \begin{pmatrix} a+b & 0 \\ 0 & b+c \end{pmatrix}$.

a) Write down the domain and the codomain of T.

b) Find $T(1 + 2x + 3x^2)$ and $T(3 + 7x - 8x^2)$.

c) Does T eventually produce all of $M_{22}(\mathbb{R})$ by transforming all of $P_2(\mathbb{R})$?

a) Domain: $P_2(\mathbb{R})$ CoDomain: $M_{22}(\mathbb{R})$

b) $T(1+2x+3x^2) = \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}$

$T(3+7x-8x^2) = \begin{pmatrix} 10 & 0 \\ 0 & -1 \end{pmatrix}$

c) No, we only get

$$\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \in M_{22}(\mathbb{R})$$

We will <u>never</u> produce $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

★

Once again observe above how naturally T changes quadratics into matrices!

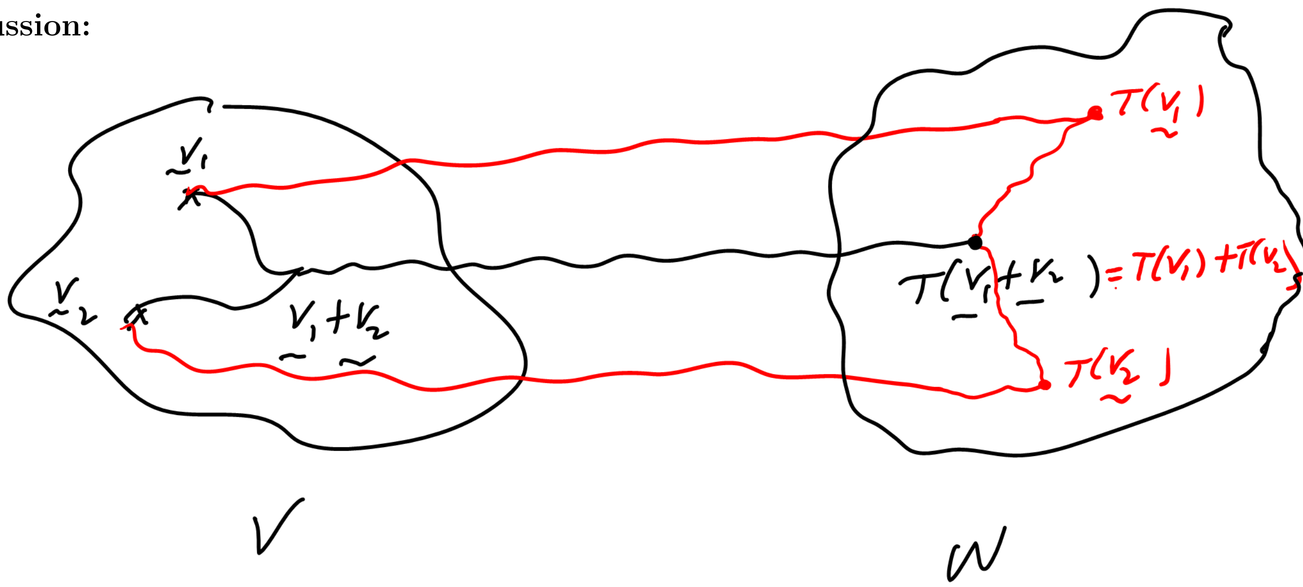
Whenever we have transformations between spaces, it is natural to demand that the transformations respect the underlying structure of the spaces being changed. Vector spaces are all about addition and scalar multiplication, so we expect the transformations between vector spaces to behave well with respect to these operations. To ensure this, we force our transformations to be **linear.**

Definition: A transformation T mapping the vector space V into the vector space W is linear if:

I) $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ and

II) $T(\alpha\mathbf{v}_1) = \alpha T(\mathbf{v}_1)$ for all scalars α (usually $\alpha \in \mathbb{R}$) and vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$.

Discussion:



★

Condition I) above states that adding vectors in V and then transforming them is the same as transforming them and then adding them in W. Similarly for the second condition.

Not all transformations are linear, however in Math1231, we will only be interested in linear transformations. You must be able to prove that given transformations are linear.

Do not confuse the concept of **linearity** with that of **closure**. The ideas look similar but are completely different! Lets take a very careful look at the proof of linearity, using the two transformations we have already considered above.

Example 3: Let the transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2y \\ 3x \\ x+y \end{pmatrix}.$$

Prove that T is linear.

We always start a proof of linearity by assigning $\mathbf{v}_1$ and $\mathbf{v}_2$ to be two arbitrary vectors in the domain of T, and α to be a scalar, usually just a real number.

(I) Let $\underset{\sim}{v_1}, \underset{\sim}{v_2} \in \mathbb{R}^2$ (choose two vectors in the domain)

$\underset{\sim}{v_1} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$, $\underset{\sim}{v_2} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ (describe them)

$$T(\underset{\sim}{v_1} + \underset{\sim}{v_2}) = T\begin{pmatrix} x_1+x_2 \\ y_1+y_2 \end{pmatrix} = \begin{pmatrix} 2(y_1+y_2) \\ 3(x_1+x_2) \\ x_1+x_2+y_1+y_2 \end{pmatrix}$$

T(sum)

$$= \begin{pmatrix} 2y_1+2y_2 \\ 3x_1+3x_2 \\ x_1+x_2+y_1+y_2 \end{pmatrix}$$

$$T(\underset{\sim}{v_1}) + T(\underset{\sim}{v_2}) = T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2y_1 \\ 3x_1 \\ x_1+y_1 \end{pmatrix} + \begin{pmatrix} 2y_2 \\ 3x_2 \\ x_2+y_2 \end{pmatrix}$$

sum of T( )'s

$$= \begin{pmatrix} 2y_1+2y_2 \\ 3x_1+3x_2 \\ x_1+y_1+x_2+y_2 \end{pmatrix}$$

$$= T(\underset{\sim}{v_1} + \underset{\sim}{v_2})$$

II) For $\alpha \in \mathbb{R}$ choose a scalar

$$T(\alpha \underset{\sim}{v_1}) = T\begin{pmatrix} \alpha x_1 \\ \alpha y_1 \end{pmatrix} = \begin{pmatrix} 3\alpha y_1 \\ 2\alpha x_1 \\ \alpha x_1 + \alpha y_1 \end{pmatrix}$$

$$\alpha T(\underset{\sim}{v_1}) = \alpha T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \alpha \begin{pmatrix} 3y_1 \\ 2x_1 \\ x_1+y_1 \end{pmatrix} = \begin{pmatrix} 3\alpha y_1 \\ 2\alpha x_1 \\ \alpha x_1 + \alpha y_1 \end{pmatrix} \overset{\star}{=} T(\alpha \underset{\sim}{v_1})$$

I & II $\Rightarrow$ T is a linear transformation

As with closure, it is much easier to prove that a transformation is **not** linear than to prove that it **is** linear. All we have to do is show any specific thing going wrong.

Example 4: Define $T : \mathbb{R}^2 \to M_{33}(\mathbb{R})$ by $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & y^2 \end{pmatrix}$.

a) Find $T\begin{pmatrix} 4 \\ 7 \end{pmatrix}$ and $T\begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

b) Prove that T is not linear.

---

a) $T\begin{pmatrix} 4 \\ 7 \end{pmatrix} = \begin{pmatrix} 16 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 49 \end{pmatrix}$

$T\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 9 \end{pmatrix}$

Observe how efficiently T transforms real two dimensional vectors into 3×3 matrices!

b) To verify non-linearity all we need is a specific example displaying one of the linearity conditions failing. We do not seek a general argument.

b) Let $\underset{\sim}{v_1} = \begin{pmatrix} 4 \\ 7 \end{pmatrix}$, $\underset{\sim}{v_2} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

$$T(\underset{\sim}{v_1} + \underset{\sim}{v_2}) = T\begin{pmatrix} 6 \\ 10 \end{pmatrix} = \begin{pmatrix} 36 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 100 \end{pmatrix}$$

$$T(v_1) + T(v_2) = \begin{pmatrix} 16 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 49 \end{pmatrix} + \begin{pmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

$$= \begin{pmatrix} 20 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 58 \end{pmatrix} \neq T(\underset{\sim}{v_1} + \underset{\sim}{v_2})$$

★

Note that non-linear transformations can be interesting but in this course we will only deal with the linear case since we then have perfect control of the situation.

Example 5: Let $T : V \to W$ be a linear transformation. Prove that $T(\mathbf{0}) = \mathbf{0}$.

$$T(\underset{\sim}{0}) = T(\underset{\sim}{0}+\underset{\sim}{0}) = T(\underset{\sim}{0}) + T(\underset{\sim}{0}) \quad \text{(linearity)}$$

$$\Rightarrow T(\underset{\sim}{0}) + T(\underset{\sim}{0}) = T(\underset{\sim}{0})$$

$$\Rightarrow T(\underset{\sim}{0}) = \underset{\sim}{0}.$$

★

Every linear transformation maps the zero vector in the domain to the zero vector in the codomain. If this doesn't happen, the transformation is **NOT** linear!

Example 6: Define $T : \mathbb{R}^5 \to \mathbb{R}^3$ by

$$T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} x_5 \\ x_3 \\ 1 \end{pmatrix}.$$

$\in \mathbb{R}^5$

a) Calculate $T\begin{pmatrix} 3 \\ 7 \\ 8 \\ 1 \\ 6 \end{pmatrix}$. $= \begin{pmatrix} 6 \\ 8 \\ 1 \end{pmatrix} \in \mathbb{R}^3$

b) Is T a linear transformation?

$$T\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \neq \underset{\sim}{0}.$$

$T(\underset{\sim}{0}) \neq \underset{\sim}{0}$ $\therefore$ T is not linear.

★

So $T(\mathbf{0}) \neq \mathbf{0}$ tells you that T is not linear,

BUT

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & y^2 \end{pmatrix}$$

$T(\mathbf{0}) = \mathbf{0}$ tells you nothing!

Example 4 above has $T\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ but we saw that the transformation T is **not** linear.

A proof of linearity is always a long formal process.

To prove non-linearity, just give a particular example of **anything** not working as it should.

To finish off, let's look at an abstract linearity proof.

Example 7: Define $T : P_2(\mathbb{R}) \to M_{22}(\mathbb{R})$ by $T(a + bx + cx^2) = \begin{pmatrix} a+b & 0 \\ 0 & b+c \end{pmatrix}$.

Prove that T is a linear transformation.

As usual, we pick off and fully describe two general vectors in the domain of T:

(I) Let $\underset{\sim}{v_1}, \underset{\sim}{v_2} \in P_2(\mathbb{R})$ (pick some vectors in domain)

$\underset{\sim}{v_1} = a_1 + b_1 x + c_1 x^2$, $\underset{\sim}{v_2} = a_2 + b_2 x + c_2 x^2$ (description)

$$T(\underset{\sim}{v_1} + \underset{\sim}{v_2}) = T\left((a_1 + a_2) + (b_1 + b_2)x + (c_1 + c_2)x^2\right)$$

T(sum)

$$= \begin{pmatrix} a_1 + a_2 + b_1 + b_2 & 0 \\ 0 & b_1 + b_2 + c_1 + c_2 \end{pmatrix}$$

$$T(\underset{\sim}{v_1}) + T(\underset{\sim}{v_2}) = T(a_1 + b_1 x + c_1 x^2) + T(a_2 + b_2 x + c_2 x^2)$$

sum of T( )'s

$$= \begin{pmatrix} a_1 + b_1 & 0 \\ 0 & b_1 + c_1 \end{pmatrix} + \begin{pmatrix} a_2 + b_2 & 0 \\ 0 & b_2 + c_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_1 + b_1 + a_2 + b_2 & 0 \\ 0 & b_1 + c_1 + b_2 + c_2 \end{pmatrix} = T(\underline{v_1} + \underline{v_2})$$

II) Let $\alpha \in \mathbb{R}$.

$$T(\alpha \underset{\sim}{v_1}) = T(\alpha a_1 + \alpha b_1 x + \alpha c_1 x^2) = \begin{pmatrix} \alpha a_1 + \alpha b_1 & 0 \\ 0 & \alpha b_1 + \alpha c_1 \end{pmatrix}$$

$$\alpha T(\underset{\sim}{v_1}) = \alpha \begin{pmatrix} a_1 + b_1 & 0 \\ 0 & b_1 + c_1 \end{pmatrix} = \begin{pmatrix} \alpha a_1 + \alpha b_1 & 0 \\ 0 & \alpha b_1 + \alpha c_1 \end{pmatrix}$$

$$= T(\alpha \underset{\sim}{v_1})$$

I ⊕ II $\Rightarrow$ T is $\underline{\underline{\text{linear}}}$ ★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 9

LINEAR TRANSFORMATIONS ON CONCRETE SPACES

Milan Pahor

MATH1231 ALGEBRA

LINEAR TRANSFORMATIONS ON CONCRETE SPACES

A transformation T from the vector space V into the vector space W is linear if the following two conditions are satisfied:

$$(I) \qquad T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$$

and

$$(II) \qquad T(\alpha \mathbf{v}_1) = \alpha T(\mathbf{v}_1)$$

for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ and scalars α.

FACT: Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation. Suppose that $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$. Then the $m \times n$ matrix A which implements the transformation T is given by

$$A = \left(T(\mathbf{e}_1) \vdots T(\mathbf{e}_2) \vdots \cdots \vdots T(\mathbf{e}_n) \right)$$

That is, A is simply the vertical assemblage of all the transformed standard basis vectors, and pre-multiplying a vector by A is the same as transforming the vector.

In the last lecture we looked at linear transformations $T : V \to W$ where V and W were **any** vector spaces. In this lecture we will focus exclusively on the situation where the vector spaces are instead $\mathbb{R}^n$ and $\mathbb{R}^m$ and $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$. In this simpler situation we have absolute and complete control and once again Gaussian elimination comes into play.

Example 1: Suppose that a linear transformation $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^4$ is such that

$$T\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad T\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}.$$

a) Find $T\begin{pmatrix} -1 \\ 4 \end{pmatrix}$.

b) Find $T\begin{pmatrix} x \\ y \end{pmatrix}$.

A remarkable feature of linear transformations is that scant knowledge of what T does to a basis immediately expands to knowledge about the entire space. This is the power of linearity!

We know what T does to $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 5 \end{pmatrix}$, so let's build $\begin{pmatrix} -1 \\ 4 \end{pmatrix}$ from these two special vectors.

$$\alpha\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \beta\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \alpha + 2\beta \\ 2\alpha + 5\beta \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$$

$$\Rightarrow \begin{matrix} \alpha + 2\beta = -1 \\ 2\alpha + 5\beta = 4 \end{matrix}$$

$$\Rightarrow \left(\begin{array}{cc|c} 1 & 2 & -1 \\ 2 & 5 & 4 \end{array}\right)$$

a) $\left(\begin{array}{cc|c} 1 & 2 & -1 \\ 2 & 5 & 4 \end{array}\right)$ ✓ ✓

$R_2 = R_2 - 2R_1$ $\left(\begin{array}{cc|c} 1 & 2 & -1 \\ 0 & 1 & 6 \end{array}\right)$

$\beta = 6$

$\alpha + 2\beta = -1$

$\alpha + 12 = -1$

$\alpha = -13$

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

$$T\begin{pmatrix} -1 \\ 4 \end{pmatrix} = T\left(-13\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 6\begin{pmatrix} 2 \\ 5 \end{pmatrix}\right)$$

(But T is linear !!)

$$= -13\,T\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 6\,T\begin{pmatrix} 2 \\ 5 \end{pmatrix} = -13\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + 6\begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} -13 \\ 0 \\ -13 \\ 0 \end{pmatrix} + \begin{pmatrix} 18 \\ 6 \\ 6 \\ -6 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ -7 \\ -6 \end{pmatrix}$$

b) Now let's build a random $\begin{pmatrix} x \\ y \end{pmatrix}$ from the two special vectors:

$$\left(\begin{array}{cc|c} 1 & 2 & x \\ 2 & 5 & y \end{array}\right) \qquad R_2 = R_2 - 2R_1 \qquad \left(\begin{array}{cc|c} 1 & 2 & x \\ 0 & 1 & y-2x \end{array}\right)$$

$$\therefore \ \beta = y - 2x.$$

$$\alpha + 2\beta = x \Rightarrow \alpha + 2(y-2x) = x$$

$$\alpha + 2y - 4x = x$$

$$\alpha = 5x - 2y$$

$$\therefore \ \begin{pmatrix} x \\ y \end{pmatrix} = (5x-2y)\begin{pmatrix} 1 \\ 2 \end{pmatrix} + (y-2x)\begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

$$T\begin{pmatrix} x \\ y \end{pmatrix} = T\left[(5x-2y)\begin{pmatrix} 1 \\ 2 \end{pmatrix} + (y-2x)\begin{pmatrix} 2 \\ 5 \end{pmatrix}\right]$$

$$= (5x-2y)\,T\begin{pmatrix} 1 \\ 2 \end{pmatrix} + (y-2x)\,T\begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

$$= (5x-2y)\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + (y-2x)\begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

$$= \begin{pmatrix} 5x-2y+3y-6x \\ y-2x \\ 5x-2y+y-2x \\ 2x-y \end{pmatrix} = \begin{pmatrix} y-x \\ y-2x \\ 3x-y \\ 2x-y \end{pmatrix}$$

$$\star \quad a) \ \ T\begin{pmatrix} -1 \\ 4 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ -7 \\ -6 \end{pmatrix} \qquad b) \ \ T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y-x \\ y-2x \\ 3x-y \\ 2x-y \end{pmatrix} \quad \star$$

Let's check the formula in b) by evaluating $T\begin{pmatrix} 2 \\ 5 \end{pmatrix}$

$$T\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 5-2 \\ 5-4 \\ 6-5 \\ 4-5 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix} \checkmark\checkmark$$

$\star$

Example 2: Suppose that a transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ is such that

$$T\begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}1\\5\end{pmatrix}, \quad T\begin{pmatrix}3\\1\\2\end{pmatrix} = \begin{pmatrix}4\\3\end{pmatrix} \text{ and } T\begin{pmatrix}-5\\1\\-6\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}$$

Prove that T is not linear.

Let's see if we can write $\begin{pmatrix}-5\\1\\-6\end{pmatrix}$ as a linear combination of the other two vectors in $\mathbb{R}^3$.

$$\left(\begin{array}{cc|c}1 & 3 & -5\\1 & 1 & 1\\0 & 2 & -6\end{array}\right) \xRightarrow{R_2 = R_2 - R_1} \left(\begin{array}{cc|c}1 & 3 & -5\\0 & -2 & 6\\0 & 2 & -6\end{array}\right)$$

$$R_3 = R_3 + R_2 \quad \left(\begin{array}{cc|c}1 & 3 & -5\\0 & -2 & 6\\0 & 0 & 0\end{array}\right)$$

$-2\beta = 6 \Rightarrow \beta = -3$

$\alpha + 3\beta = -5$

$\alpha - 9 = -5 \Rightarrow \alpha = 4$

$$\therefore \begin{pmatrix}-5\\1\\-6\end{pmatrix} = 4\begin{pmatrix}1\\1\\0\end{pmatrix} - 3\begin{pmatrix}3\\1\\2\end{pmatrix}$$

$$T\begin{pmatrix}-5\\1\\-6\end{pmatrix} = T\left(4\begin{pmatrix}1\\1\\0\end{pmatrix} - 3\begin{pmatrix}3\\1\\2\end{pmatrix}\right)$$

$$= 4T\begin{pmatrix}1\\1\\0\end{pmatrix} - 3T\begin{pmatrix}3\\1\\2\end{pmatrix} = 4\begin{pmatrix}1\\5\end{pmatrix} - 3\begin{pmatrix}4\\3\end{pmatrix}$$

$$= \begin{pmatrix}4\\20\end{pmatrix} - \begin{pmatrix}12\\9\end{pmatrix} = \begin{pmatrix}-8\\11\end{pmatrix} \neq \begin{pmatrix}1\\1\end{pmatrix}$$

$\therefore$ T is not linear. ★

Discussion: Say that instead we had $T\begin{pmatrix}-5\\1\\-6\end{pmatrix} = \begin{pmatrix}-8\\11\end{pmatrix}$. What would that mean?

Nothing

★ Nothing, you cannot prove linearity by looking at specific vectors. It might just be luck. ★

MATRICES AS LINEAR TRANSFORMATIONS

A truly remarkable fact is that **every** $m \times n$ matrix serves as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ through the simple process of matrix multiplication.

Example 3: Let A be the 5×2 matrix

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 2 & 1 \\ 3 & 5 \\ 0 & 1 \end{pmatrix}$$

$(5\times 2)(2\times 1) = (5\times 1)$

and define a transformation T by $T(\mathbf{v}) = A\mathbf{v}$ where $A\mathbf{v}$ is just A multiplied by $\mathbf{v}$ on the right. Note that for linear transformations we **always** place the vector on the right of the matrix.

$\mathbb{R}^2$ $\mathbb{R}^5$

a) Find the domain and codomain of T. $T: \mathbb{R}^2 \to \mathbb{R}^5$

b) Find $T\begin{pmatrix} 2 \\ 6 \end{pmatrix}$.

c) Prove that T is linear.

b) $T\begin{pmatrix} 2 \\ 6 \end{pmatrix} = A\begin{pmatrix} 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 2 & 1 \\ 3 & 5 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 6 \end{pmatrix}$

$(5\times 2)\times(2\times 1) = (5\times 1)$

$= \begin{pmatrix} 14 \\ 2 \\ 10 \\ 36 \\ 6 \end{pmatrix}$

$(m\times n): \mathbb{R}^n \to \mathbb{R}^m$

$T: \mathbb{R}^2 \to \mathbb{R}^5$

★

Note above that matrix multiplication is intrinsically a linear process, so if a transformation is defined through a matrix it is easy to verify linearity. As an alternative we could also argue that

$$T\begin{pmatrix} x \\ y \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 2 & 1 \\ 3 & 5 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+2y \\ x \\ 2x+y \\ 3x+5y \\ y \end{pmatrix}$$

and then use the formal approach from the last lecture to prove linearity directly from the definition. Let's do it!

Let $\underset{\sim}{v_1}, \underset{\sim}{v_2} \in \mathbb{R}^2$ $\qquad T(\underset{\sim}{v}) = A\underset{\sim}{v}$

$$\begin{aligned} T(\underset{\sim}{v_1} + \underset{\sim}{v_2}) &= A(\underset{\sim}{v_1} + \underset{\sim}{v_2}) \\ &= A\underset{\sim}{v_1} + A\underset{\sim}{v_2} \\ &= T(\underset{\sim}{v_1}) + T(\underset{\sim}{v_2}) \end{aligned}$$

For scalar α.

$$T(\alpha \underset{\sim}{v_1}) = A(\alpha \underset{\sim}{v_1}) = \alpha A \underset{\sim}{v_1} = \alpha T(\underset{\sim}{v_1})$$

$\therefore$ T is linear.

Example 4: Example 3 re-done using the formal definition of linearity:

Define $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^5$ by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 2 & 1 \\ 3 & 5 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+2y \\ x \\ 2x+y \\ 3x+5y \\ y \end{pmatrix}.$$

(I) Prove **from the definition** that $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$.

Let $\underset{\sim}{v_1}, \underset{\sim}{v_2} \in \mathbb{R}^2$

$\underset{\sim}{v_1} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \underset{\sim}{v_2} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$

Then

$$T(\underset{\sim}{v_1} + \underset{\sim}{v_2}) = T\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right)$$

$$T\begin{pmatrix} x_1+x_2 \\ y_1+y_2 \end{pmatrix} = \begin{pmatrix} x_1+x_2+2y_1+2y_1 \\ x_1+x_2 \\ 2x_1+2x_2+y_1+y_2 \\ 3x_1+3x_2+5y_1+5y_2 \\ y_1+y_2 \end{pmatrix}$$

$$T(\underset{\sim}{v_1}) + T(\underset{\sim}{v_2}) = T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1+2y_1 \\ x_1 \\ 2x_1+y_1 \\ 3x_1+5y_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2+2y_2 \\ x_2 \\ 2x_2+y_2 \\ 3x_2+5y_2 \\ y_2 \end{pmatrix}$$

$$= \begin{pmatrix} x_1+2y_1+x_2+2y_2 \\ x_1+x_2 \\ 2x_1+y_1+2x_2+y_2 \\ 3x_1+5y_1+3x_2+5y_2 \\ y_1+y_2 \end{pmatrix}$$

$$= T(\underset{\sim}{v_1} + \underset{\sim}{v_2})$$

(II) **Homework:**

Prove that $T(\alpha\mathbf{v}_1) = \alpha T(\mathbf{v}_1)$ for all $\mathbf{v}_1 \in \mathbb{R}^2$ and for all scalars $\alpha \in \mathbb{R}$.

★

Discussion: Say I wish to transform $\mathbb{R}^{17}$ into $\mathbb{R}^{33}$. What dimension matrix do I need for the job?

★

We have seen above that **EVERY** $m \times n$ matrix can, if we wish, be viewed as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ with the transformation being implemented via the process of matrix multiplication.

An absolutely amazing fact is that the opposite is also true!. That is, **EVERY** linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be viewed as being implemented by matrix multiplication for some $m \times n$ matrix. How do we find the matrix? We have two ways:

Example 5: Define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2y \\ 3x \\ x+y \end{pmatrix}.$$

a) Evaluate $T\begin{pmatrix} 4 \\ 1 \end{pmatrix}$.

b) Find a matrix A so that $T(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^2$.
(A is often called a matrix repesentation of T)

c) Check your answer by evaluating the matrix product $A\begin{pmatrix} 4 \\ 1 \end{pmatrix}$.

---

a) $T\begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 12 \\ 5 \end{pmatrix}$

b) **Method 1:** Before we even start we know that the matrix will need to be 3×2 to get from $\mathbb{R}^2$ to $\mathbb{R}^3$.

Let's fill out the transformation

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0x+2y \\ 3x+0y \\ 1x+1y \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 3 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Clearly $A = \begin{pmatrix} 0 & 2 \\ 3 & 0 \\ 1 & 1 \end{pmatrix}$ and as a check $\begin{pmatrix} 0 & 2 \\ 3 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 12 \\ 5 \end{pmatrix}$

$(3 \times 2) \times (2 \times 1) = (3 \times 1)$

b) **Method 2:** The second method, though formal, is of enormous use when dealing with geometric transformations for which there is no actual formula.

FACT: Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation. Suppose that $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$. Then the $m \times n$ matrix A which implements T is given by

$$A = \left(T(\mathbf{e}_1) \vdots T(\mathbf{e}_2) \vdots \cdots \vdots T(\mathbf{e}_n) \right)$$

That is, A is simply the vertical assemblage of all the transformed standard basis vectors, and pre-multiplying a vector by A is the same as transforming the vector.

So:

$$T : \mathbb{R}^2 \to \mathbb{R}^3$$

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2y \\ 3x \\ x+y \end{pmatrix}$$

Basis for $\mathbb{R}^2$ $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$

$$T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}, \quad T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 2 \\ 3 & 0 \\ 1 & 1 \end{pmatrix}$$

$$\left(T(e_1) \vdots T(e_2) \right)$$

★

Using Method 2 we can now actually produce matrices which DO things to space!

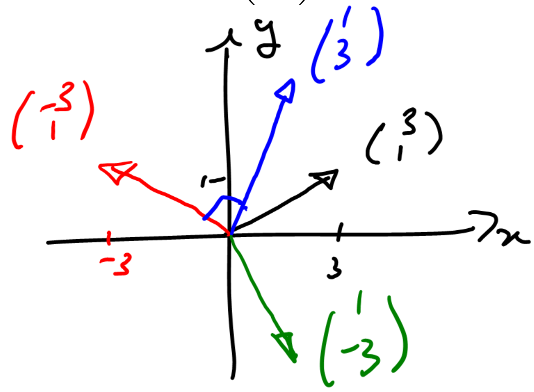
Example 6: Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which first reflects vectors in $\mathbb{R}^2$ in the y axis , then rotates them clockwise by 90^o and then finally reflects them in the x axis. The matrix A of T will be 2×2.....but which 2×2 matrix do we need?

a) Produce $T\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ manually.

b) Find the matrix representation A of T by considering the action of T on the standard basis $B = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$.

c) Check your matrix A by pre-multiplying $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ by A and comparing with a).

a) $T\begin{pmatrix} 3 \\ 1 \end{pmatrix}$

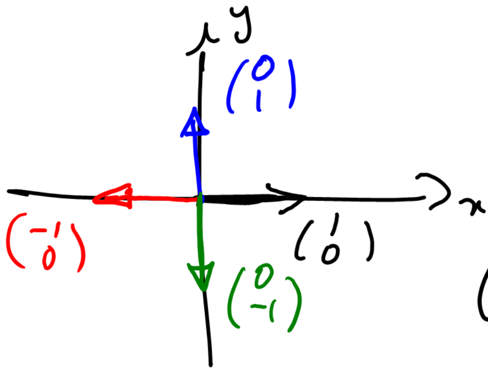


$\begin{pmatrix} -3 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 3 \end{pmatrix} = 0.$

$\therefore T\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$

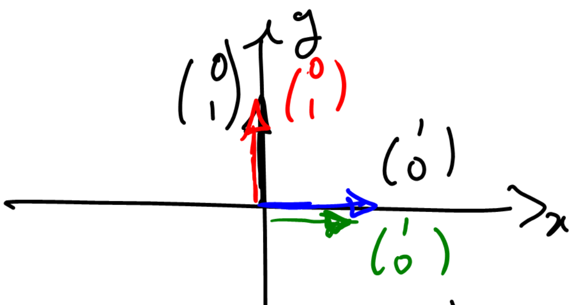
So we have $\begin{pmatrix} 3 \\ 1 \end{pmatrix} \xrightarrow{\text{Reflect } y\text{-axis}} \begin{pmatrix} -3 \\ 1 \end{pmatrix} \xrightarrow{\text{Rotate } \circlearrowright 90^\circ} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \xrightarrow{\text{Reflect } x\text{-axis}} \begin{pmatrix} 1 \\ -3 \end{pmatrix}$

b) $T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$



$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \begin{pmatrix} -1 \\ 0 \end{pmatrix} \to \begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

$T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$



$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

So $A = \left(T(e_1) \vdots T(e_2) \right) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$

check $T\begin{pmatrix} 3 \\ 1 \end{pmatrix} = A\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$ ✓✓

A does <u>all</u> of these things <u>at once</u>

★ $a)$ $\begin{pmatrix} 1 \\ -3 \end{pmatrix}$ $b)$ $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ ★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 10

THE RANK NULLITY THEOREM

Milan Pahor

MATH1231 ALGEBRA
THE RANK NULLITY THEOREM

FACT: Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear transformation. Suppose that $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is the standard basis for $\mathbb{R}^n$. Then the $m \times n$ matrix A which implements the transformation T is given by

$$A = \Big(T(\mathbf{e}_1) \vdots T(\mathbf{e}_2) \vdots \cdots \vdots T(\mathbf{e}_n) \Big)$$

That is, A is simply the vertical assemblage of all the transformed standard basis vectors, and pre-multiplying a vector by A is the same as transforming the vector.

Suppose that $T : V \to W$ is a linear map where V and W are vector spaces. Then

$$\mathrm{Ker}(T) = \{\mathbf{v} \in V \,|\, T(\mathbf{v}) = 0\}. \text{ (The kernel of T)}$$

$$\mathrm{Im}(T) = \{T(\mathbf{v}) \,|\, \mathbf{v} \in V\}. \text{ (The image of T)}$$

Ker(T) is a subspace of V and Im(T) is a subspace of W.

Dim(Ker(T)) is referred to as the nullity of T, and denoted by Nullity(T).

Dim(Im(T)) is referred to as the rank of T, and denoted by Rank(T).

The Rank-Nullity theorem states that

$$\mathrm{Rank}(T){+}\mathrm{Nullity}(T){=}\mathrm{Dim}(V).$$

All the Rank-Nullity theorem says is that "what you destroy plus what you produce equals what you started with".

In this lecture we will explore the Rank-Nullity Theorem, a central result in the theory of linear transformations. But first another example of the creation of the standard matrix A of a particular linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

Example 1: Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which projects all vectors in $\mathbb{R}^2$ onto the vector $\begin{pmatrix} 9 \\ 3 \end{pmatrix}$.

a) Calculate $T\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ and display the situation with a diagram.

b) Find the standard matrix P of T by considering the action of T on the standard basis $B = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$.

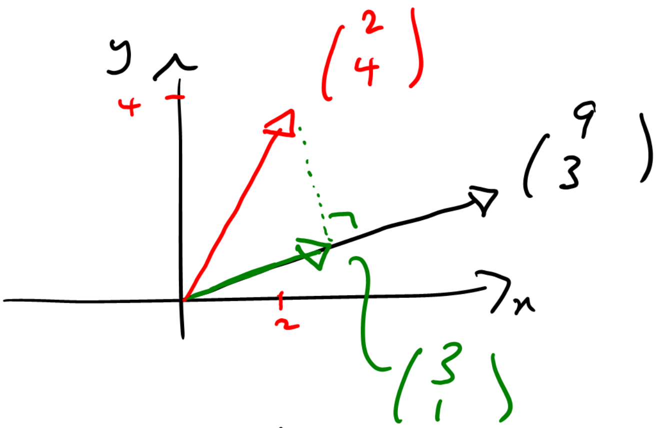
c) Calculate the matrix product $P\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ and compare with a).

d) Evaluate P^2. Explain the significance of your answer.

a) $$T\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \mathrm{Proj}_{\begin{pmatrix} 9 \\ 3 \end{pmatrix}}\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \frac{\begin{pmatrix} 2 \\ 4 \end{pmatrix}\cdot\begin{pmatrix} 9 \\ 3 \end{pmatrix}}{\begin{pmatrix} 9 \\ 3 \end{pmatrix}\cdot\begin{pmatrix} 9 \\ 3 \end{pmatrix}}\begin{pmatrix} 9 \\ 3 \end{pmatrix}$$

$$= \frac{30}{90}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$\therefore\ T\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$



b) $$T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\cdot\begin{pmatrix} 9 \\ 3 \end{pmatrix}}{\begin{pmatrix} 9 \\ 3 \end{pmatrix}\cdot\begin{pmatrix} 9 \\ 3 \end{pmatrix}}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \frac{9}{90}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{9}{10} \\ \frac{3}{10} \end{pmatrix}$$

$$T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mathrm{Proj}_{\begin{pmatrix} 9 \\ 3 \end{pmatrix}}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\cdot\begin{pmatrix} 9 \\ 3 \end{pmatrix}}{\begin{pmatrix} 9 \\ 3 \end{pmatrix}\cdot\begin{pmatrix} 9 \\ 3 \end{pmatrix}}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \frac{3}{90}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \frac{1}{30}\begin{pmatrix} 9 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{3}{10} \\ \frac{1}{10} \end{pmatrix}$$

$$\therefore\ P = \begin{pmatrix} \frac{9}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} \end{pmatrix}$$

$$\therefore \quad T(\underset{\sim}{v}) = P\underset{\sim}{v}$$

check $$T\begin{pmatrix}2\\4\end{pmatrix} = \begin{pmatrix}\frac{9}{10} & \frac{3}{10}\\ \frac{3}{10} & \frac{1}{10}\end{pmatrix}\begin{pmatrix}2\\4\end{pmatrix}$$

$$\underbrace{(2\times 2)\times(2\times 1)} = (2\times 1)$$

$$= \begin{pmatrix}\frac{18}{10} + \frac{12}{10}\\ \frac{6}{10} + \frac{4}{10}\end{pmatrix} = \begin{pmatrix}3\\1\end{pmatrix}$$

as before !!

c) $$P^2 = \begin{pmatrix}\frac{9}{10} & \frac{3}{10}\\ \frac{3}{10} & \frac{1}{10}\end{pmatrix}\begin{pmatrix}\frac{9}{10} & \frac{3}{10}\\ \frac{3}{10} & \frac{1}{10}\end{pmatrix}$$

$$= \begin{pmatrix}\frac{81}{100} + \frac{9}{100} & \frac{27}{100} + \frac{3}{100}\\ \frac{27}{100} + \frac{3}{100} & \frac{9}{100} + \frac{1}{100}\end{pmatrix}$$

$$= \begin{pmatrix}\frac{9}{10} & \frac{3}{10}\\ \frac{3}{10} & \frac{1}{10}\end{pmatrix} \quad !!$$

$$\underline{P^2 = P}$$

$P^2 = P \times P \Rightarrow$ Projecting Twice !!

Projecting Twice $\equiv$ Projecting Once !

$$\bigstar \quad P = \begin{pmatrix}\frac{9}{10} & \frac{3}{10}\\ \frac{3}{10} & \frac{1}{10}\end{pmatrix} \quad \bigstar$$

Every linear transformation $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ can be viewed as matrix multiplication by some 2×2 matrix. The following is a list of some examples of transformations and their matrices. You do not need to memorise these, but they may be handy in the future.

SOME STANDARD MATRIX TRANSFORMATIONS ON $\mathbb{R}^2$

Reflection in the x axis: $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

Reflection in the y axis: $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

Reflection across the origin: $A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$

Rotation anticlockwise about the origin by θ: $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

For clockwise rotations use $-\theta$ in the above matrix.

Dilation by α in the x direction and β in the y direction: $A = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$

$\alpha, \beta > 1$ yields expansions $\quad \alpha, \beta < 1$ yields contractions.

$k-$shear in the x direction: $A = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$

$k-$shear in the y direction: $A = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$

Reflection in the line $y = mx$: $A = \dfrac{1}{1+m^2} \begin{pmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{pmatrix}$

Projection onto the line $y = mx$: $A = \dfrac{1}{1+m^2} \begin{pmatrix} 1 & m \\ m & m^2 \end{pmatrix}$

In all of the above the inverse of the transformation (if it exists) is given by A^{-1}.

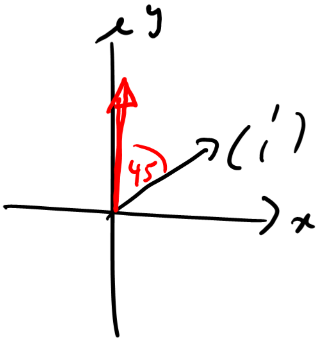
Example 2:

a) Use the previous table to find the matrix R which rotates vectors in $\mathbb{R}^2$ anti-clockwise by 45^o about the origin.

b) Have a look at $R\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and see if this makes sense.

c) Evaluate R^8 without making any calculations.

a) $$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos 45 & -\sin 45 \\ \sin 45 & \cos 45 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$



b) $$R\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{2}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix}$$

c) $R^8 = I$

Rotating 8x by 45° is the same as doing nothing!

$$\star \quad R = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \quad \star$$

So we have projection of $\mathbb{R}^2$ onto the vector $\begin{pmatrix} 9 \\ 3 \end{pmatrix}$ implemented by $P = \begin{pmatrix} \frac{9}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} \end{pmatrix}$

and rotation anti-clockwise by 45^o implemented by $R = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$

Our final task is to consider the composition of transformations, that is following one transformation up with another. Remarkably this is achieved by simply multiplying the standard matrices **in the correct order**. It is for this reason that matrix multiplication is actually defined in such a bizarre way.

Example 3: The linear transformation T firstly projects onto the vector $\begin{pmatrix} 9 \\ 3 \end{pmatrix}$ and then rotates anticlockwise by 45°. Find its standard matrix A.

$$A = RP = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{9}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} \end{pmatrix} = \begin{pmatrix} \frac{6}{10\sqrt{2}} & \frac{2}{10\sqrt{2}} \\ \frac{12}{10\sqrt{2}} & \frac{4}{10\sqrt{2}} \end{pmatrix}$$ ~ projects first then rotates

$A = R(P\,\underset{\sim}{v})$

★

It feels that the order of R and P is wrong above but it is right. The vector being transformed goes on the right, so the first matrix it encounters needs to be P.

Example 4: The linear transformation T first rotates anticlockwise by 45° and then projects onto the vector $\begin{pmatrix} 9 \\ 3 \end{pmatrix}$. Find its standard matrix A.

$$A = PR = \begin{pmatrix} \frac{9}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{12}{10\sqrt{2}} & \frac{-6}{10\sqrt{2}} \\ \frac{4}{10\sqrt{2}} & \frac{-2}{10\sqrt{2}} \end{pmatrix}$$ ~ rotates first then projects.

★

Observe that the matrices are completely different. It matters which transformation goes first!

Example 3 explains why matrix multiplication has such a strange definition. Matrix multiplication needs to organise the composition of transformations.

It should also now be clear why matrix multiplication is non-commutative, that is $RP \neq PR$. It really does matter what you do first.

The Rank-Nullity Theorem

Linear transformations often act between spaces of differing dimension. As a result they can be quite violent, completely annihilating big chunks of the domain V. If the transformation $T : V \to W$ is linear we call the bits of V killed by T the kernel of T, denoted by $\mathrm{Ker}(T)$. Whatever is ultimately produced over in the space W is referred to as the image of T, denoted by $\mathrm{Im}(T)$. That is:

Suppose that $T : V \to W$ is a linear map where V and W are vector spaces. Then

$$\mathrm{Ker}(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = 0\}. \text{ (The kernel of T)}$$
$$\mathrm{Im}(T) = \{T(\mathbf{v}) \mid \mathbf{v} \in V\}. \text{ (The image of T)}$$

$\mathrm{Ker}(T)$ is a subspace of V and $\mathrm{Im}(T)$ is a subspace of W.

$\mathrm{Dim}(\mathrm{Ker}(T))$ is referred to as the nullity of T, and denoted by $\mathrm{Nullity}(T)$.

$\mathrm{Dim}(\mathrm{Im}(T))$ is referred to as the rank of T, and denoted by $\mathrm{Rank}(T)$.

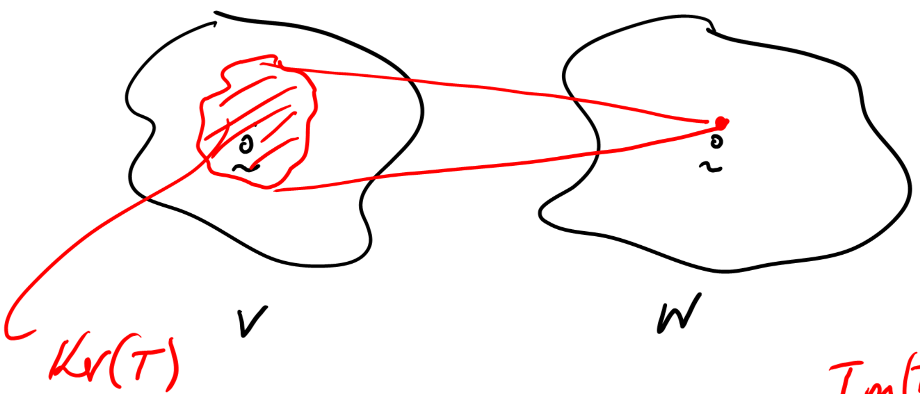
The Rank-Nullity theorem states that

killed

produced $\mathrm{Rank}(T)+\mathrm{Nullity}(T)=\mathrm{Dim}(V)$. ~ started with.

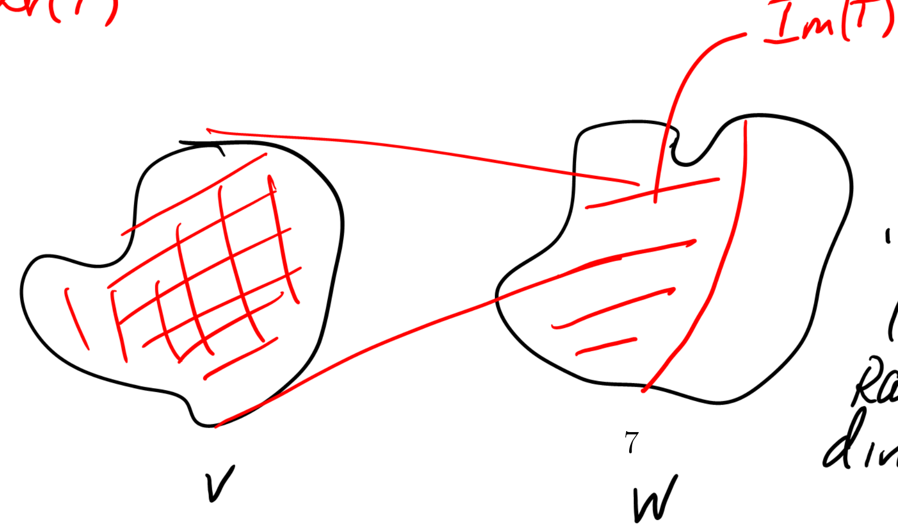
All the Rank-Nullity theorem says is that "what you destroy plus what you produce equals what you started with".

Discussion: $T : V \to W$



Ker(T) is the subspace of V "killed" by T

Nullity(T) = Dim(Ker(T))



Im(T) is the "subspace of W produced" by T

Rank(T) is the dimension of Im(T).

Example 5: Define a linear transformation $T : P_2(\mathbb{R}) \to \mathbb{R}^4$ by

$$T(a + bx + cx^2) = \begin{pmatrix} a+b \\ 0 \\ b+c \\ 0 \end{pmatrix}.$$

a) Write down the domain and codomain of T together with their respective dimensions.

b) Evaluate $T(3 + 4x + 6x^2)$.

c) Find $\mathrm{Im}(T)$, a basis for $\mathrm{Im}(T)$ and $\mathrm{Rank}(T)$.

d) Find $\mathrm{Ker}(T)$, a basis for $\mathrm{Ker}(T)$ and $\mathrm{Nullity}(T)$.

e) Check that T annihilates the basis vectors from d).

f) Verify the Rank-Nullity theorem for this linear transformation.

a) $\mathrm{Dom}(T) = P_2(\mathbb{R})$ $\quad$ Dim $= 3$

$\quad$ $\mathrm{Co\text{-}Dom}(T) = \mathbb{R}^4$ $\quad$ Dim $= 4$

b) $T(3+4x+6x^2) = \begin{pmatrix} 7 \\ 0 \\ 10 \\ 0 \end{pmatrix}$

c) By inspection: Basis $(\mathrm{Im}(T)) = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right\}$

$$\begin{pmatrix} a+b \\ 0 \\ b+c \\ 0 \end{pmatrix} = (a+b)\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + (b+c)\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

$\mathrm{Rank}(T) = 2$

d) $\underline{\mathrm{Ker}(T)\ ??}$

Assume $T(a+bx+cx^2) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} a+b \\ 0 \\ b+c \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\therefore \quad \begin{aligned} a+b &= 0 \\ b+c &= 0 \end{aligned}$$

So $b = -c$

$a = -b = --c = c.$

$$\therefore \ a + bx + cx^2 = c - cx + cx^2$$
$$= c(1 - x + x^2)$$

$\therefore$ Basis $\mathrm{Ker}(T) = \{1 - x + x^2\}$

$$T(1-x+x^2) = \begin{pmatrix} 1-1 \\ 0 \\ -1+1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \checkmark\checkmark$$

$\therefore$ Nullity $(T) = 1$

$$\begin{aligned} \mathrm{Rank}(T) + \mathrm{Nullity}(T) &= 2 + 1 \\ &= 3 \\ &= \mathrm{Dim}(P_2(\mathbb{R})) \end{aligned}$$

$\therefore$ Verified.

★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 11

RANK-NULLITY THEOREM CONTINUED

Milan Pahor

MATH1231 ALGEBRA
RANK-NULLITY THEOREM CONTINUED

Suppose that $T : V \to W$ is a linear map where V and W are vector spaces. Then

$$\text{Ker}(T) = \{\mathbf{v} \in V \,|\, T(\mathbf{v}) = 0\}. \text{ (The kernel of T)}$$

$$\text{Im}(T) = \{T(\mathbf{v}) \,|\, \mathbf{v} \in V\}. \text{ (The image of T)}$$

Ker(T) is a subspace of V and Im(T) is a subspace of W.

Dim(Ker(T)) is referred to as the nullity of T, and denoted by Nullity(T).

Dim(Im(T)) is referred to as the rank of T, and denoted by Rank(T).

The Rank-Nullity theorem states that

$$\text{Rank}(T)+\text{Nullity}(T)=\text{Dim}(V).$$

All the Rank-Nullity theorem says is that "what you destroy plus what you produce equals what you started with".

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation given by multiplication by a matrix A. That is $T\mathbf{v} = A\mathbf{v}$ for some $m \times n$ matrix A.

Then, after reducing A to echelon form:

- Ker(T) is simply the solution of $A\mathbf{x} = \mathbf{0}$.
- Nullity(T) is the number of **non-leading** columns in the echelon form.
- Im(T)$\equiv$col(A) is the span of all the leading columns of A.
- Rank(T) is the number of **leading** columns in the Echelon form.
- All the Rank-Nullity theorem says is, that in the echelon form of the matrix A:

The number of leading columns+the number of non-leading columns=the number of columns.

In this lecture we will explore the Rank-Nullity Theorem as it applies to the special case where $T : \mathbb{R}^n \to \mathbb{R}^m$ and the action of T is described through multiplication by a particular matrix A. Once again you will see that we have immaculate control in this situation and that **all** issues are resolved by appealing to the echelon form.

Recall that if $T : \mathbb{R}^n \to \mathbb{R}^m$, the kernel of T is all the unfortunate vectors in $\mathbb{R}^n$ that get wiped out by T. The dimension of the kernel of T is called the nullity of T.

The image of T is the subspace of $\mathbb{R}^m$ which holds all of the vectors produced by T. The dimension of the image of T is called the rank of T.

It can easily be shown (see your printed notes) that $\mathrm{Ker}(T)$ and $\mathrm{Im}(T)$ are not just subsets of the domain and codomain respectively, they are in fact **subspaces**.

We do have the ability to easily answer any Rank-Nullity question when $T : \mathbb{R}^n \to \mathbb{R}^m$. Let's explore the theory through a big example. The approach will be summarised after the example.

Look very carefully at this problem and observe that when a linear transformation is expressed as pre-multiplication by a matrix A, the properties of the echelon form of A completely describe the nature of the transformation.

$\mathbb{R}^5 \quad \mathbb{R}^3$

Example 1: Let $A = \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 1 & 2 & 2 & 6 & 5 \\ 1 & 2 & 3 & 9 & 6 \end{pmatrix}$ and define $T : \mathbb{R}^n \to \mathbb{R}^m$ to be the linear transformation $T(\mathbf{v}) = A\mathbf{v}$.

$T: \mathbb{R}^5 \to \mathbb{R}^3$

a) Find n and m. $n=5, \; m=3$

b) Find $T\begin{pmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 2 \end{pmatrix}$

c) By considering $T\begin{pmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 2 \end{pmatrix}$ explain why $\mathrm{Im}(T)=\mathrm{span}\{\text{columns of } A\}=\mathrm{Col}(A)$.

d) Explain why T is linear.

e) Find $T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$.

f) Find a basis for $\mathrm{Ker}(T)$ and $\mathrm{Nullity}(T)$.

g) Check your basis in f) by transforming one of the basis vectors.

h) Find a basis for $\mathrm{Im}(T)$ and $\mathrm{Rank}(T)$.

i) Verify the Rank-Nullity theorem for this transformation.

a) Since the matrix is 3×5 we have $n = 5$ and $m = 3$ and hence $T : \mathbb{R}^5 \to \mathbb{R}^3$.

b) $$T\begin{pmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 1 & 2 & 2 & 6 & 5 \\ 1 & 2 & 3 & 9 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 14 \\ 13 \\ 12 \end{pmatrix}.$$

c) $$T\begin{pmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 1 & 2 & 2 & 6 & 5 \\ 1 & 2 & 3 & 9 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 2 \end{pmatrix}$$

$$= 1\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 4\begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} + 0\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + (-1)\begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix} + 2\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 14 \\ 13 \\ 12 \end{pmatrix}.$$

Observe that the vector $\begin{pmatrix} 14 \\ 13 \\ 12 \end{pmatrix}$ in Im(T) is simply a linear combination of the columns of A. This is always the case! Thus Im(T)=span{columns of A}=Col(A).

d) Matrix Multiplication is linear.

e) $$T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 1 & 2 & 2 & 6 & 5 \\ 1 & 2 & 3 & 9 & 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 + 2x_2 + x_3 + 3x_4 + 4x_5 \\ x_1 + 2x_2 + 2x_3 + 6x_4 + 5x_5 \\ x_1 + 2x_2 + 3x_3 + 9x_4 + 6x_5. \end{pmatrix}$$

$T : \mathbb{R}^5 \to \mathbb{R}^3$

f) $$\begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 1 & 2 & 2 & 6 & 5 \\ 1 & 2 & 3 & 9 & 6 \end{pmatrix} \xRightarrow[R_3 = R_3 - R_1]{R_2 = R_2 - R_1} \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 2 & 6 & 2 \end{pmatrix}$$

$R_3 = R_3 - 2R_2$

$$\begin{array}{c} \begin{array}{ccccc} \checkmark & \times & \checkmark & \times & \times \end{array} \\ \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$Ker(T)$: Solve $\quad T(\underset{\sim}{v}) = \underset{\sim}{0} \iff A\underset{\sim}{v} = \underset{\sim}{0}$

$$\begin{array}{c} \begin{array}{ccccc} \checkmark & \times & \checkmark & \times & \times \end{array} \\ \left(\begin{array}{ccccc|c} 1 & 2 & 1 & 3 & 4 & 0 \\ 0 & 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \end{array}$$

Let $\quad x_2 = \mu, \; x_4 = \lambda, \; x_5 = t$

$x_3 + 3x_4 + x_5 = 0 \to x_3 + 3\lambda + t = 0$

$x_3 = -3\lambda - t$

$x_1 + 2x_2 + x_3 + 3x_4 + 4x_5 = 0 \Rightarrow x_1 + 2\mu + (-3\lambda - t) + 3\lambda + 4t = 0$

$x_1 + 2\mu \cancel{-3\lambda} - t \cancel{+3\lambda} + 4t = 0 \Rightarrow x_1 = -2\mu - 3t$

$$Ker(T) : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\mu + \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix}\lambda + \begin{pmatrix} -3 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}t$$

$$Basis(Ker(T)) = \left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}, \; Nullity = 3$$

g) check $\quad T\begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 3 & 4 \\ 1 & 2 & 2 & 6 & 5 \\ 1 & 2 & 3 & 9 & 6 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

$(3 \times 5) \times (5 \times 1) = (3 \times 1)$

h) $Basis(Im(T)) = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}, \; Rank(T) = 2$

$T : \mathbb{R}^5 \to \mathbb{R}^3$

$Rank(T) + Nullity(T) = Dim(V)$

$2 + 3 = 5$

(# leading cols) + (# non-leading cols) = # cols in A.

$$\bigstar\, e)\, T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + x_3 + 3x_4 + 4x_5 \\ x_1 + 2x_2 + 2x_3 + 6x_4 + 5x_5 \\ x_1 + 2x_2 + 3x_3 + 9x_4 + 6x_5 \end{pmatrix} \quad f)\, \mathrm{Ker}(T) = \mathrm{span}\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\} \quad \bigstar$$

$$\bigstar \quad h) \quad \mathrm{Im}(T) = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\} \qquad \mathrm{Rank}(T) = 2 \quad \bigstar$$

$$\bigstar \quad i) \quad \mathrm{Rank}(T) + \mathrm{Nullity}(T) = 2 + 3 = 5 = \mathrm{Dim}(\mathbb{R}^5) \quad \bigstar$$

Observe in h) above, that the Rank Nullity theorem for a matrix transformation simply states that the number of leading columns, $\mathrm{Rank}(T)$, plus the number of non leading columns, $\mathrm{Nullity}(T)$, is equal to the number of columns, $\mathrm{Dim}(V)$.

Note that we now have a complete description of the action of T. It maps $\mathbb{R}^5$ to $\mathbb{R}^3$,

annihilating the directions $\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \right\}$ in $\mathbb{R}^5$ and producing the

plane spanned by $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}$ in $\mathbb{R}^3$.

Summary

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation given by pre-multiplication by a matrix A. That is $T\mathbf{v} = A\mathbf{v}$ for some $m \times n$ matrix A.

Then, after reducing A to echelon form:

- $\mathrm{Ker}(T)$ is simply the solution of $A\mathbf{x} = \mathbf{0}$.

- $\mathrm{Nullity}(T)$ is the number of **non-leading** columns in the echelon form.

- $\mathrm{Im}(T) \equiv \mathrm{col}(A)$ is the span of all the leading columns of A.

- $\mathrm{Rank}(T)$ is the number of **leading** columns in the Echelon form.

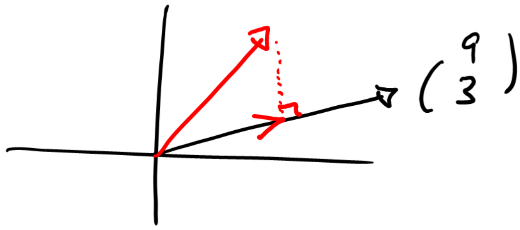
- All the Rank-Nullity theorem says is, that in the echelon form of the matrix A:

The number of leading columns+the number of non-leading columns=the number of columns.

Example 2: Recall in example 1 in the previous lecture, we considered the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ which projects all vectors in $\mathbb{R}^2$ onto the vector $\begin{pmatrix} 9 \\ 3 \end{pmatrix}$.

We showed that $T(\mathbf{v}) = P\mathbf{v}$ where

$$P = \begin{pmatrix} \frac{9}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} \end{pmatrix}.$$



a) Find Ker(T), Nullity(T), Im(T) and Rank(T) by considering the geometry of T.

b) Find Ker(T), Nullity(T), Im(T) and Rank(T) by considering the echelon form of P.

a)

$\mathrm{Ker}(T)$: $\mathrm{span}\left\{\begin{pmatrix} -3 \\ 9 \end{pmatrix}\right\}$

note $\begin{pmatrix} -3 \\ 9 \end{pmatrix} \cdot \begin{pmatrix} 9 \\ 3 \end{pmatrix} = -27 + 27 = 0$

$\therefore \begin{pmatrix} -3 \\ 9 \end{pmatrix} \perp \begin{pmatrix} 9 \\ 3 \end{pmatrix}$

Nullity $= 1$

$\mathrm{Im}(T) = \mathrm{span}\left\{\begin{pmatrix} 9 \\ 3 \end{pmatrix}\right\}$

Rank$(T) = 1$

Rank Nullity Th^m

$1 + 1 = 2$

b)

$\begin{pmatrix} \frac{9}{10} & \frac{3}{10} \\ \frac{3}{10} & \frac{1}{10} \end{pmatrix} \xrightarrow[R_2 = 10R_2]{R_1 = 10R_1} \begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix} \xrightarrow{R_2 = R_2 - \frac{1}{3}R_1} \begin{pmatrix} 9 & 3 \\ 0 & 0 \end{pmatrix}$

✓ ✗

$\mathrm{Ker}(P)$: $\left(\begin{array}{cc|c} 9 & 3 & 0 \\ 0 & 0 & 0 \end{array}\right)$

Let $x_2 = t$

$9x_1 + 3t = 0 \Rightarrow 9x_1 = -3t$

$x_1 = -\frac{1}{3}t \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} \\ 1 \end{pmatrix} t$

$\mathrm{Ker}(P) = \mathrm{span}\left\{\begin{pmatrix} -\frac{1}{3} \\ 1 \end{pmatrix}\right\} = \mathrm{span}\left\{\begin{pmatrix} -1 \\ 3 \end{pmatrix}\right\}$

$= \mathrm{span}\left\{\begin{pmatrix} -3 \\ 9 \end{pmatrix}\right\}$

Nullity $= 1$

$\mathrm{Basis}(\mathrm{Im}(P)) = \left\{\begin{pmatrix} \frac{9}{10} \\ \frac{3}{10} \end{pmatrix}\right\}$

$\equiv \left\{\begin{pmatrix} 9 \\ 3 \end{pmatrix}\right\}$ ★

Rank$(P) = 1$

It is clear from the above analysis that when transformations are defined through matrix multiplication, the calculation of Ker(T), Nullity(T), Im(T) and Rank(T) are easily made. But when a transformation is defined only as a *process*, the situation demands a more hands-on approach .

Example 3: Define $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ by

$$T(p) = p'.$$

a) Calculate $T(3 + 5x + 4x^2)$.

b) Prove that T is linear. (This is NOT the usual big calculation!)

c) What is the dimension of the domain of T?

d) Find a basis for Im(T) and hence Rank(T).

e) Find a basis for Ker(T) and hence Nullity(T).

f) Verify the Rank-Nullity theorem in this case.

a) $T(3+5x+4x^2)$

$= 5+8x.$

b) Differentiation is a linear process.

c) $\mathrm{Dim}(P_2(\mathbb{R})) = 3$

d) Basis $(\mathrm{Im}(T))$

$=\{1, x\}$

$\therefore$ Rank $(T) = 2$

e) Ker (T) = all constant functions!

$= \mathrm{span}\{1\}$

Nullity $T = 1$

Rank (T) + Nullity (T) = Dim (V)

$2 + 1 = 3$

★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 12

EIGENVALUES AND EIGENVECTORS

Milan Pahor

MATH1231 ALGEBRA
EIGENVALUES AND EIGENVECTORS

Given a square matrix A, a non-zero vector $\mathbf{v}$ is said to be an eigenvector of A if

$$A\mathbf{v} = \lambda\mathbf{v}$$

for some $\lambda \in \mathbb{R}$. The number λ is referred to as the associated eigenvalue of A.

To complete an eigenanalysis, first find the eigenvalues through the characteristic equation $\det(A - \lambda I) = 0$. The eigenvectors are then found via row reduction and back substitution.

Every eigenvalue has attached to it, infinitely many different dependent eigenvectors. We usually just choose one that looks nice.

The zero vector is **never** an eigenvector but it is OK to have a zero eigenvalue.

Ideally an $n \times n$ matrix has n linearly independent eigenvectors.......but you can run short! The only general way to find out if a matrix has a full set of eigenvectors is to find them all.

A useful check is the fact that $\Sigma(\text{eigenvalues}) = \text{Trace}(A)$.

Eigenvectors from different eigenvalues are linearly independent.

Establishing the eigenanalysis of a matrix gives you a clear vision of the internal workings of that particular matrix.

This lecture will introduce the subtle and quite complicated theory of eigenvalues and eigenvectors for a square matrix A.

But first a little revision on determinants.

Example 1: Find the determinant of each of the following matrices and find the inverse if it exists. Check your answer for the inverse.

a) $E = \begin{pmatrix} 3 & 8 \\ 2 & 6 \end{pmatrix}$ b) $F = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 2 & 4 \end{pmatrix}$.

Recall the following crucial facts regarding inverses and determinants.

- The inverse of a matrix A is another matrix A^{-1} with the property that $AA^{-1} = I$.
- $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \dfrac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$
- $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$
- Larger inverses and determinants are best calculated using Gaussian elimination or Maple.
- A^{-1} exists $\Longleftrightarrow \det(A) \neq 0$.
- $A\mathbf{v} = 0$ has non-trivial solutions for $\mathbf{v} \Longleftrightarrow A^{-1}$ **does not** exist $\Longleftrightarrow \det(A) = 0$.

---

a) $\det\begin{pmatrix} 3 & 8 \\ 2 & 6 \end{pmatrix}$

$= \begin{vmatrix} 3 & 8 \\ 2 & 6 \end{vmatrix}$

$= 3 \times 6 - 2 \times 8$

$= 18 - 16 = 2$

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \dfrac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

$\begin{pmatrix} 3 & 8 \\ 2 & 6 \end{pmatrix}^{-1} = \dfrac{1}{2}\begin{pmatrix} 6 & -8 \\ -2 & 3 \end{pmatrix}$

$= \begin{pmatrix} 3 & -4 \\ -1 & 3/2 \end{pmatrix} = A^{-1}$

check $\begin{pmatrix} 3 & 8 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} 3 & -4 \\ -1 & 3/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$= I$

$\therefore$ ✓✓

b) $\begin{vmatrix} \overset{+}{1} & \overset{-}{2} & \overset{+}{3} \\ 1 & 0 & 1 \\ 2 & 2 & 4 \end{vmatrix}$

$1\begin{vmatrix} 0 & 1 \\ 2 & 4 \end{vmatrix} - 2\begin{vmatrix} 1 & 1 \\ 2 & 4 \end{vmatrix} + 3\begin{vmatrix} 1 & 0 \\ 2 & 2 \end{vmatrix}$

$= 1(0-2) - 2(4-2) + 3(2-0)$

$= -2 - 4 + 6 = \underline{0}$

$\therefore \det = 0$

$\therefore$ inverse fails to exist.

★

Eigenvalues and Eigenvectors

Consider the matrix $\begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix}$ and have a look at what A does to a random vector:

$\begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix} \begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 16 \\ 15 \end{pmatrix}$......it's nothing special!

But now consider $\begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 21 \end{pmatrix} = 7 \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

Observe that A simply makes this special vector 7 times as long!

We say that $\mathbf{v} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is an eigenvector of A with associated eigenvalue $\lambda = 7$.

How do we find all the eigenvectors and eigenvalues of a matrix A? Well

$$A\mathbf{v} = \lambda\mathbf{v} \Longrightarrow A\mathbf{v} = \lambda I\mathbf{v} \Longrightarrow A\mathbf{v} - \lambda I\mathbf{v} = 0 \Longrightarrow (A - \lambda I)\mathbf{v} = \mathbf{0}.$$

Now $\mathbf{v} = 0$ is the trivial solution to the above matrix equation and we are seeking non-trivial solutions. Thus the matrix $A - \lambda I$ must be non-invertible and hence we demand that

$$\boxed{\det(A - \lambda I) = 0.}$$

This is called the characteristic equation and generates the eigenvalues. 2×2 matrices have a quadratic characteristic equation and 3×3 matrices will have a cubic characteristic equation. Once you have solved the characteristic equation to find all the eigenvalues, you can then find the eigenvectors by solving $(A - \lambda I)\mathbf{v} = \mathbf{0}$ using row reduction.

It is a long and involved process! Fortunately there are lots of checks along the way!

$$A\underset{\sim}{v} = \lambda \underset{\sim}{v}$$

Example 2: Find all the eigenvalues and eigenvectors of $A = \begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix}$, checking your calculations at every opportunity.

$$A - \lambda I = \begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-\lambda & 4 \\ -3 & 9-\lambda \end{pmatrix}$$

$$\det(A - \lambda I) = \det \begin{pmatrix} 1-\lambda & 4 \\ -3 & 9-\lambda \end{pmatrix} = 0$$

$$(1-\lambda)(9-\lambda) - -12 = 9 - \lambda - 9\lambda + \lambda^2 + 12 = 0$$

$$\Rightarrow \lambda^2 - 10\lambda + 21 = 0$$

$$\Rightarrow (\lambda - 3)(\lambda - 7) = 0$$

$$\therefore \lambda = \underline{\underline{3, 7}} \quad \text{(Eigenvalues!!)}$$

<u>check</u>: Σ e-vals $= \Sigma$ diag elements of A

$= \text{trace}(A)$

$3 + 7 = 1 + 9$ ✓✓

<u>Eigenvectors</u>

$$A\underset{\sim}{v} = \lambda \underset{\sim}{v}$$

<u>$\lambda = 3$</u>: $A\underset{\sim}{v} = 3\underset{\sim}{v} \Rightarrow (A - 3\underline{I})(\underset{\sim}{v}) = \underset{\sim}{0}$

$$\left(\begin{array}{cc|c} 1-3 & 4 & 0 \\ -3 & 9-3 & 0 \end{array}\right) \Rightarrow \left(\begin{array}{cc|c} -2 & 4 & 0 \\ -3 & 6 & 0 \end{array}\right)$$

✓ ×

$R_2 = R_2 - \frac{3}{2} R_1$

$$\left(\begin{array}{cc|c} -2 & 4 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

must happen !!

Let $y = t$

$-2x + 4t = 0 \Rightarrow 2x = 4t \Rightarrow x = 2t$

$$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2t \\ t \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} t, \quad t \in \mathbb{R}$$

In particular $\underset{\sim}{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \Longleftarrow$ eigenvector

$\lambda = 3$, e-vecs $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 200 \\ 100 \end{pmatrix}, \begin{pmatrix} 6 \\ 3 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \end{pmatrix}$

∞ many of them.

eigenspace $E_3 = \text{span}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}$

check $\begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 3 \end{pmatrix} = 3\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$\therefore A\underset{\sim}{v} = 3\underset{\sim}{v}$

$\lambda = 7$

$$\left(\begin{array}{cc|c} 1-7 & 4 & 0 \\ -3 & 9-7 & 0 \end{array}\right)$$

$$\left(\begin{array}{cc|c} -6 & 4 & 0 \\ -3 & 2 & 0 \end{array}\right)$$

$R_2 = R_2 - \frac{1}{2}R_1$

$$\left(\begin{array}{cc|c} -6 & 4 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Let $y = t$

$-6x + 4t = 0 \Rightarrow 6x = 4t$

$x = \frac{4}{6}t = \frac{2}{3}t$

$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2/3 \\ 1 \end{pmatrix} t$

In particular $\underset{\sim}{v} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

Note that **any** non-zero multiple of an eigenvector is again an eigenvector.

★ Eigenvalues are $\lambda = 3$ and 7 with associated independent eigenvectors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ & $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ resp. ★

★ We can also say that the eigenspaces are $E_3 = \text{span}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}$ and $E_7 = \text{span}\left\{\begin{pmatrix} 2 \\ 3 \end{pmatrix}\right\}$. ★

★ A basis for $\mathbb{R}^2$ consisting of eigenvectors of A is thus $\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}\right\}$ ★

check $\begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 21 \end{pmatrix} = 7\begin{pmatrix} 2 \\ 3 \end{pmatrix}$

$\therefore A\underset{\sim}{v} = 7\underset{\sim}{v}$

Example 3: Find all the eigenvalues and eigenvectors of $B = \begin{pmatrix} 7 & 1 \\ -1 & 9 \end{pmatrix}$, checking your calculations at every opportunity.

$$\begin{vmatrix} 7-\lambda & 1 \\ -1 & 9-\lambda \end{vmatrix} = 0$$

$$\Rightarrow (7-\lambda)(9-\lambda) + 1 = 0$$

$$\Rightarrow 63 - 7\lambda - 9\lambda + \lambda^2 + 1 = 0$$

$$\Rightarrow \lambda^2 - 16\lambda + 64 = 0$$

$$\Rightarrow (\lambda - 8)^2 = 0$$

$$\therefore \lambda = 8, 8.$$

check
Σ e-vals $= Tr(B)$
$8+8 = 9+7$ ✓✓

$A\underset{\sim}{v} = 8\underset{\sim}{v}$
$\therefore (A - 8I)\underset{\sim}{v} = \underset{\sim}{0}$

$\underline{\lambda = 8}$: $\left(\begin{array}{cc|c} -1 & 1 & 0 \\ -1 & 1 & 0 \end{array}\right)$ ✓ ✗

$R_2 = R_2 - R_1$ $\left(\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$

Let $y = t$
$-x + t = 0$
$x = t$

$$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t \\ t \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} t$$

In particular $\underset{\sim}{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector for eigenvalue $\lambda = 8$.

$$\begin{pmatrix} 7 & 1 \\ -1 & 9 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 8 \end{pmatrix} = 8\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\therefore A\underset{\sim}{v} = \lambda\underset{\sim}{v}.$$

$\bigstar$ Eigenvalues are $\lambda = 8$ and 8 with (only one!) eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. $\bigstar$

We say that B has a single one dimensional eigenspace $E_8 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$.

Discussion: Is there a basis for $\mathbb{R}^2$ consisting of eigenvectors of B? NO

Discussion: How about $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 5 \end{pmatrix} \right\}$ as a basis of eigenvectors for $\mathbb{R}^2$?

dependent !! $\therefore$ Not a basis

What is happening with B is that we do not have a full set of eigenvectors. Any self-respecting 2×2 matrix really **should** have two linearly independent eigenvectors.

Although rare, a shortage of eigenvectors is always very,very bad news for the matrix, and generally we try to avoid such delinquent matrices.

Note that if a matrix has its eigenvalues all different, then there is no chance of being short eigenvectors. Everything will be OK. Even when some eigenvalues repeat, there is still a chance that everything will work properly!

Example 4: Find all the eigenvalues and eigenvectors of $C = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}$.

In this course we will tend to focus on 2×2 matrices, as anything bigger leads to horrendous calculations, and is best done via Maple.

The matrix C is an interesting example where we are short eigenvalues, but still end up with a full set of eigenvectors! It is impossible to see this coming. You just have to make a cup of tea, sit down and find all the eigenvectors. If there is enough then there is enough.

We start with the characteristic polynomial $\det(A - \lambda I) = 0$. If at all possible we will try to avoid the situation where we actually produce a cubic polynomial equation as these are difficult to solve.

$$\begin{vmatrix} -2-\lambda & 2 & -3 \\ 2 & 1-\lambda & -6 \\ -1 & -2 & -\lambda \end{vmatrix} = (-2-\lambda)\begin{vmatrix} 1-\lambda & -6 \\ -2 & -\lambda \end{vmatrix} - 2\begin{vmatrix} 2 & -6 \\ -1 & -\lambda \end{vmatrix} - 3\begin{vmatrix} 2 & 1-\lambda \\ -1 & -2 \end{vmatrix}$$

$$= (-2-\lambda)\{-\lambda(1-\lambda) - 12\} - 2\{-2\lambda - 6\} - 3\{-4 + 1 - \lambda\}$$

$$= (-2-\lambda)\{\lambda^2 - \lambda - 12\} - 2\{-2\lambda - 6\} - 3\{-3 - \lambda\}$$

$$= (-2-\lambda)\{\lambda^2 - \lambda - 12\} + 4\lambda + 12 + 9 + 3\lambda$$

$$= (-2-\lambda)(\lambda - 4)(\lambda + 3) + 7\lambda + 21$$

$$= (-2-\lambda)(\lambda - 4)(\lambda + 3) + 7(\lambda + 3)$$

$$= (\lambda + 3)\{(-2-\lambda)(\lambda - 4) + 7\}$$

$$= (\lambda + 3)\{-\lambda^2 + 2\lambda + 15\}$$

$$= -(\lambda + 3)\{\lambda^2 - 2\lambda - 15\}$$

$$= -(\lambda + 3)(\lambda + 3)(\lambda - 5) = 0.$$

check.
Σ e-vals $= Tr(A)$
$-3 + -3 + 5 = -2 + 1 + 0$
$-1 = -1$ ✓✓

Thus $\lambda = -3,\ -3,\ 5$.

As a check $\quad -3 + -3 + 5 = -2 + 1 + 0.$

Note that the fact that $\lambda = -3$ has doubled up is certainly troubling but it does not imply that we necessarily will be short an eigenvector. Let's now find the eigenvectors, first for $\lambda = -3$:

$A - \lambda I$

$$\left(\begin{array}{ccc|c} -2+3 & 2 & -3 & 0 \\ 2 & 1+3 & -6 & 0 \\ -1 & -2 & 0+3 & 0 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & 0 \\ 2 & 4 & -6 & 0 \\ -1 & -2 & 3 & 0 \end{array}\right) \xrightarrow[R_3 = R_3 + R_1]{R_2 = R_2 - 2R_1} \begin{array}{c} \checkmark \quad \times \quad \times \\ \left(\begin{array}{ccc|c} 1 & 2 & -3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \end{array}$$

Let $y = \mu,\ z = \lambda$

$$x + 2y - 3z = 0 \implies x + 2\mu - 3\lambda = 0$$
$$\implies x = -2\mu + 3\lambda$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}\mu + \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}\lambda$$

$\therefore$ e-vecs for $\lambda = -3$ are $\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} \right\}$

$$E_{-3} = \operatorname{span}\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} \right\}$$

eigenspace <u>is</u> a <u>plane</u> !!

check $\begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -9 \\ 0 \\ -3 \end{pmatrix} = -3\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$

$$\therefore\ A\underset{\sim}{v} = -3\underset{\sim}{v} \quad \checkmark$$

Homework: Find the remaining eigenvector for $\lambda = 5$.

$$\left(\begin{array}{ccc|c} -2-5 & 2 & -3 & 0 \\ 2 & 1-5 & -6 & 0 \\ -1 & -2 & 0-5 & 0 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|c} -7 & 2 & -3 & 0 \\ 2 & -4 & -6 & 0 \\ -1 & -2 & -5 & 0 \end{array}\right) \xrightarrow[R_1 = -R_1]{R_1 \Leftrightarrow R_3}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 2 & -4 & -6 & 0 \\ -7 & 2 & -3 & 0 \end{array}\right) \xrightarrow[R_3 = R_3 + 7R_1]{R_2 = R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 0 & -8 & -16 & 0 \\ 0 & 16 & 32 & 0 \end{array}\right) \xrightarrow[R_3 = R_3 + 2R_2]{}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 5 & 0 \\ 0 & -8 & -16 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Let $z = t$

$-8y - 16t = 0 \Rightarrow y = -2t$

$x + 2(-2t) + 5(t) = 0$

$\Rightarrow x = -t$

$\therefore \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix} t$

and in particular $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$

★ $\lambda = -3, -3$ and 5 with associated eigenvectors $\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ ★

We still ended up with a full set of eigenvectors, even though the ice was a bit thin!

So the three eigenvectors $\left\{ \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \right\}$ will form a basis for $\mathbb{R}^3$.

The only time you can be short eigenvectors is when the eigenvalues start repeating, BUT having eigenvalues repeat does not imply that you **will** run short......just that it might happen.

It is also possible for both eigenvalues and/or eigenvectors to be non-real! These complex eigenvectors are still calculated in the same way, but the algebra is even messier. This does not happen very often in Math1231, but is an unfortunate fact of life in reality.

Example 5: The 4×4 matrix $\begin{pmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{pmatrix}$ has characteristic equation

$$(\lambda - 12)(\lambda - \beta)^3 = 0.$$

Find β.

$\lambda = 12, \beta, \beta, \beta.$

Σ e-vals $= \mathrm{Tr}(A)$

$\Rightarrow 12 + 3\beta = 36 \Rightarrow 3\beta = 24 \Rightarrow \underline{\underline{\beta = 8}}$

$\therefore$ e-vals $12, 8, 8, 8$ ★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 13

DIAGONALISATION

Milan Pahor

MATH1231 ALGEBRA
DIAGONALISATION

Given a square matrix A, a non-zero vector $\mathbf{v}$ is said to be an eigenvector of A if $A\mathbf{v} = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{R}$. The number λ is referred to as the associated eigenvalue of A.

We first find eigenvalues through the characteristic equation $\det(A - \lambda I) = 0$. The eigenvectors are then found via row reduction and back substitution.

If an $n \times n$ matrix A has n linearly independent eigenvectors and P is the matrix of eigenvectors aligned vertically then $P^{-1}AP = D$ where D is the diagonal matrix of eigenvalues. The order of the eigenvalues in D must match the order of the eigenvectors in P. This is referred to as a diagonalisation of A.

A matrix is non-diagonalisable if it is short on eigenvectors. The only general way to find out if a matrix has a full set of eigenvectors is to find them all.

Establishing the eigenanalysis of a particular matrix gives you a clear vision of the internal workings of that matrix, and through diagonalisation the matrix may be transformed into a more workable diagonal structure.

In the previous lecture we discussed how to calculate eigenvectors and eigenvalues of a square matrix A. It is a gruesome task, and sometimes a full set of eigenvectors is not even available. Eigenvectors are of enormous importance in matrix theory and are used in a very subtle way.

Consider the matrix
$$D = \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix}.$$
We call such a matrix a **diagonal** matrix as it only has non-zero entries on the leading diagonal. In fact it is almost all zero! A 3×3 diagonal matrix would look like
$$D = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Diagonal matrices are the world's most boring and stable matrices. They are incredibly easy to work with and nothing ever goes wrong with a diagonal matrix.

By using eigenvalues and eigenvectors we can convert **any** matrix into a diagonal matrix in a highly controlled fashion! This process, called diagonalisation, enables us to analyse complicated matrices as if they were quite simple!

To diagonalise a matrix A we need to find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$. This converts complicated A into simple D.

Let's take a look at the process of diagonalisation using some of the matrices from the previous lecture. We will close the lecture with a proof that our technique is valid.

Example 1: Consider the square 2×2 matrix $A = \begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix}$.

Find an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

We proved in the previous lecture that

$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is an eigenvector of A with eigenvalue $\lambda = 3$.

We also showed that

$\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ is an eigenvector of A with eigenvalue $\lambda = 7$.

This is all we need! We then have:

$$P = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} \Longrightarrow P^{-1}AP = \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix} = D.$$

That is, to diagonalise A via $P^{-1}AP = D$, let P be the matrix of eigenvectors of A aligned vertically and D be the diagonal matrix of associated eigenvalues. The order of the eigenvalues in D must match the order of the eigenvectors in P.

It is certainly not obvious that this works! We will prove the result after a couple of examples.

The diagonalisation above is a bridge between the complicated $A = \begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix}$ and the simple $D = \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix}$.

The whole theory of eigenvectors is plagued by a lack of uniqueness. This is also true of the process of diagonalisation.

Example 2: Find two more different diagonalisations of the above matrix A.

$$P = \begin{pmatrix} 2 & 2 \\ 3 & 1 \end{pmatrix} \Longrightarrow P^{-1}AP = \begin{pmatrix} 7 & 0 \\ 0 & 3 \end{pmatrix} = D.$$

$$P = \begin{pmatrix} 20 & 2 \\ 30 & 1 \end{pmatrix} \Longrightarrow P^{-1}AP = \begin{pmatrix} 7 & 0 \\ 0 & 3 \end{pmatrix} = D.$$

★

Discussion: Why is it not true that $P^{-1}AP = A$?

The P^{-1} and the P do not meet!

A CHECK?

Suppose we wished to check that

$$P = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} \Longrightarrow P^{-1}AP = \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix} = D.$$

Generally a check is not a good idea as it will take some time. But if you do feel a need to check a diagonalisation, rather than checking that $P^{-1}AP = D$, it is easier to avoid inverses and check the equivalent statement $AP = PD$. So

$$AP = \begin{pmatrix} 1 & 4 \\ -3 & 9 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 6 & 14 \\ 3 & 21 \end{pmatrix}.$$

Carefully note in the above calculation that, since the columns of P are eigenvectors of A, pre-multiplication by A simply scales the columns up by the respective eigenvalues! This will be important later on in our general proof.

Also

$$PD = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix} = \begin{pmatrix} 6 & 14 \\ 3 & 21 \end{pmatrix}.$$

Observe once again that, since the columns of P are eigenvectors of A, post-multiplication by the diagonal matrix of eigenvalues D will again scale the columns up by the respective eigenvalues!

In other words **both** AP and PD will simply scale the eigenvector columns of P by the respective eigenvalues.....this is why they are equal!

$$AP = PD$$

$$\Rightarrow \quad P^{-1}AP = P^{-1}PD$$

$$P^{-1}AP = D$$

Example 3: Consider the square 3×3 matrix $C = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}$.

Find an invertible matrix P and a diagonal matrix D such that $P^{-1}CP = D$.

We proved in the previous lecture that

$\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$ are eigenvectors of C with eigenvalue $\lambda = -3$.

We also showed that

$\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ is an eigenvector of C with eigenvalue $\lambda = 5$.

So

$$P = \begin{pmatrix} 3 & -2 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & -1 \end{pmatrix} \Longrightarrow P^{-1}CP = \begin{pmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{pmatrix} = D.$$

Example 4: Find two more different diagonalisations of the above matrix C.

$$P = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \Longrightarrow P^{-1}CP = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix} = D.$$

$$P = \begin{pmatrix} \frac{1}{2} & -2 & 3 \\ 1 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \Longrightarrow P^{-1}CP = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{pmatrix} = D.$$

★

Let's work a complete example from scratch.

Example 5: Diagonalise the matrix $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$.

This sounds like a simple request but it will take a lot of work!

$$\begin{vmatrix} 8-\lambda & -10 \\ 5 & -7-\lambda \end{vmatrix} = 0$$

$(\det(A-\lambda I)=0)$

$$(8-\lambda)(-7-\lambda)+50=0$$
$$-56-8\lambda+7\lambda+\lambda^2+50=0$$
$$\lambda^2-\lambda-6=0$$
$$(\lambda-3)(\lambda+2)=0$$
$$\therefore \quad \underline{\underline{\lambda=-2,\ 3}}$$

$\underline{\underline{\text{check}}}$ $-2+3 = 8+(-7)$

$1=1$ ✓✓

$\underline{\lambda=3}$: $\left(\begin{array}{cc|c} 5 & -10 & 0 \\ 5 & -10 & 0 \end{array}\right)$

$R_2 = R_2 - R_1$ $\left(\begin{array}{cc|c} 5 & -10 & 0 \\ 0 & 0 & 0 \end{array}\right)$

Let $y=t$

$5x-10t=0$

$x=2t$

$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} t$, $t\in\mathbb{R}$, $t\neq 0$

In particular $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$\lambda=-2$ $\left(\begin{array}{cc|c} 10 & -10 & 0 \\ 5 & -5 & 0 \end{array}\right)$

$R_2 = R_2 - \frac{1}{2}R_1$ $\left(\begin{array}{cc|c} 10 & -10 & 0 \\ 0 & 0 & 0 \end{array}\right)$

Let $y=t$

$10x = 10t \Rightarrow x=t$

$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} t$

In particular $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

$\underline{\underline{\text{check}}}$ $\begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$

$= -2\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$\therefore A\underset{\sim}{v} = -2\underset{\sim}{v}$.

$P = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ then

$P^{-1}AP = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} = D$

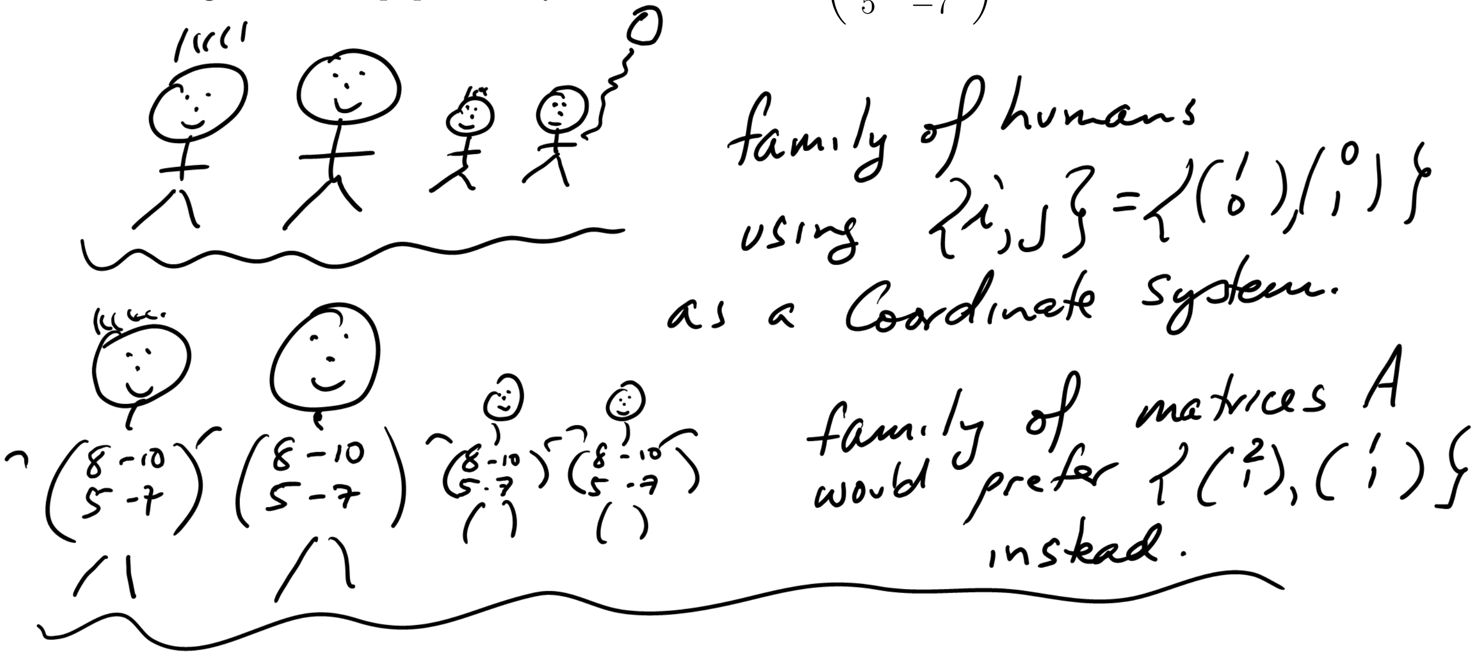
$$\bigstar \quad P = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \Longrightarrow P^{-1}AP = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} = D \quad \bigstar$$

You should not make it a habit to check that $P^{-1}AP$ is in fact equal to D, as the check is longer than the proof. If you absolutely must check something, then it is easier to check instead the equivalent statement $AP = PD$. At least that way you do not need to fiddle with inverses.

CONCLUSION

When we think of $\mathbb{R}^2$ we humans like to use $\{\mathbf{i}, \mathbf{j}\}$ as a basis. But the vectors $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ mean nothing to a particular matrix A.

Imagine a world populated by the matrices $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$.



If you were to ask one of the matrices inhabiting this universe what *it* would prefer as a basis, A would surely respond by saying

"I'll have a basis $\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ consisting of my eigenvectors thanks".

The matrix A likes its eigenvectors since the action of A upon the eigenvectors is simply contraction and elongation. Indeed it's own eigenvectors are the only special vectors for A.

If we are prepared to abandon $\{\mathbf{i}, \mathbf{j}\}$ and instead make A happy by using the coordinate system generated by its eigenvectors, then $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$ relaxes into the trivial diagonal matrix $D = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$.

That is, $P = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ transforms the complicated $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$ into the very simple $D = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$ via $P^{-1}AP = D$.

★

In the next two lectures we will look at a host of applications for the process of matrix diagonalisation. The fact that we can (sometimes) move from a complicated matrix A to a very simple diagonal matrix D can be exploited in many ways.

Not **all** matrices can be diagonalised!

Example 6: Show that the matrix $G = \begin{pmatrix} 10 & -1 \\ 9 & 4 \end{pmatrix}$ is not diagonalisable.

$$\begin{vmatrix} 10-\lambda & -1 \\ 9 & 4-\lambda \end{vmatrix} = 0 \Rightarrow (10-\lambda)(4-\lambda)+9 = 0$$

$$\Rightarrow 40 - 10\lambda - 4\lambda + \lambda^2 + 9 = 0$$

$$\Rightarrow \lambda^2 - 14\lambda + 49 = 0$$

$$\Rightarrow (\lambda-7)^2 = 0$$

$$\therefore \lambda = 7, 7$$

Could still be diagonalisable!

$\underline{\lambda = 7}$:
$$\left(\begin{array}{cc|c} 3 & -1 & 0 \\ 9 & -3 & 0 \end{array}\right) \underset{R_2 = R_2 - 3R_1}{\Longrightarrow} \left(\begin{array}{cc|c} 3 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

eigenspace is one-dimensional

$\therefore$ Not enough e-vectors

$\therefore$ Not diagonalisable! ★

Discussion: In the above example why can't we just form P from two different multiples of the one eigenvector?

$$P(\underset{\sim}{v_1} \mid 10\underset{\sim}{v_1}) \Longrightarrow P^{-1} \text{ D.N.E.}$$

★

Let's finish with a few little proofs.

Proof of Diagonalisation Procedure

Let's prove the above claims in the 3×3 case. The proof in other dimensions is similar.

Suppose that A is a 3×3 matrix with a full set of 3 linearly independent eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and associated eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$.

Let P be the matrix of eigenvectors $P = (\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3)$. Then

$$AP = A\,(\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3) = (A\mathbf{v}_1|A\mathbf{v}_2|A\mathbf{v}_3) = (\lambda_1\mathbf{v}_1|\lambda_2\mathbf{v}_2|\lambda_3\mathbf{v}_3) = (\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3)\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} = PD.$$

Thus $AP = PD$ and pre-multiplication by P^{-1} yields $P^{-1}AP = D$ as required.

★

Example 7: Suppose that A is a square matrix with eigenvector $\mathbf{v}$ and associated eigenvalue λ. If $B = A^2$, prove that $\mathbf{v}$ is also an eigenvector of B and find the associated eigenvalue.

$A\underset{\sim}{v} = \lambda \underset{\sim}{v}$

$$B\underset{\sim}{v} = A^2\underset{\sim}{v} = (A(A\underset{\sim}{v})) = A(\lambda\underset{\sim}{v})$$
$$= \lambda(A\underset{\sim}{v})$$
$$= \lambda(\lambda\underset{\sim}{v})$$
$$= \lambda^2\underset{\sim}{v}$$

$\therefore B\underset{\sim}{v} = \lambda^2\underset{\sim}{v}$ $\therefore$ $\underset{\sim}{v}$ e-vec with e-val λ^2

★

Example 8: Show that any 2×2 matrix of the form $\begin{pmatrix} \beta & \beta \\ 0 & \beta \end{pmatrix}$ where $\beta \in \mathbb{R}$, $\underline{\beta \neq 0}$ is NOT diagonalisable.

$$\begin{vmatrix} \beta-\lambda & \beta \\ 0 & \beta-\lambda \end{vmatrix} = 0 \Rightarrow (\beta-\lambda)^2 = 0$$

$\therefore \lambda = \beta, \beta.$

✗ ✓

$\underline{\lambda=\beta}$: $\left(\begin{array}{cc|c} 0 & \beta & 0 \\ 0 & 0 & 0 \end{array}\right)$

One dimensional e-space

$\therefore$ Insufficient e-vecs.

$\therefore$ Not diagonalisable. ★

$\beta y = 0 \Rightarrow y = 0$

$x = t$

$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} t$ in particular $\underset{\sim}{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 14

MATRIX POWERS

Milan Pahor

MATH1231 ALGEBRA
MATRIX POWERS

If an $n \times n$ matrix A has n linearly independent eigenvectors and P is the matrix of eigenvectors aligned vertically then $P^{-1}AP = D$ where D is the diagonal matrix of eigenvalues. The order of the eigenvalues in D must match the order of the eigenvectors in P. This is referred to as a diagonalisation of A.

A matrix is non-diagonalisable if it is short on eigenvectors.

The only general way to find out if a matrix has a full set of eigenvectors is to find them all.

$$P^{-1}AP = D \to A = PDP^{-1} \to A^n = PD^nP^{-1}.$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We turn now to two applications of the diagonalisation process. This lecture we will examine matrix powers and in the next, systems of differential equations.

There will be many matrix products in this lecture, of which I will only grind through some. The rest are for homework.

Keep in mind that if a matrix is non-diagonalisable, that is, it is short eigenvectors, then none of what we are about to do will work. The wheels fall off completely!

Keep in mind that the calculation of the this data is labour intensive, particularly when the matrices are larger than 2×2. However the fundamental concepts and techniques are the same for all dimensions and this simple little 2×2 matrix A will show us everything that we need to know!

Almost all applications of eigenvectors spring from the diagonalisation process. We always hope that an $n \times n$ matrix will have a full set of n linearly independent eigenvectors and hence be diagonalisable. If we are short eigenvectors (the first sign will be repeating eigenvalues) then the processes of this and the next lecture will not work.

Our first application of diagonalisation is the raising of a matrix to a large power. This is a complicated process as even a single matrix product is messy. It is also a crucial theory as A^n can be interpreted as n successive applications of the induced linear transformation.

Let's start with a familiar matrix we have already played with before, $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$.

Example 1: Let $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$.

a) Calculate A^2.

b) Calculate A^{100}.

---

a) $A^2 = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}\begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix} = \begin{pmatrix} 14 & -10 \\ 5 & -1 \end{pmatrix}$

$(2 \times 2) \times (2 \times 2)$ $\qquad$ 2×2

$$\star \quad \begin{pmatrix} 14 & -10 \\ 5 & -1 \end{pmatrix} \quad \star$$

Observe that you **CANNOT** evaluate A^2 just by squaring up the entries in A. You have to actually do the calculations! It is horrible, even for a 2×2 matrix.

b) Evaluating A^{100} using this neanderthal approach will take all day! We have a wonderful short cut, which will not only calculate A^{100} but will in fact produce A^n for any integer n.

Claim: Suppose that a matrix A is diagonalised via $P^{-1}AP = D$. Then

$$A^n = PD^nP^{-1}.$$

Proof:

$$P^{-1}AP = D \Rightarrow P(P^{-1}AP) = PD$$
$$\Rightarrow AP = PD$$
$$\Rightarrow (AP)(P^{-1}) = (PD)P^{-1}$$
$$\Rightarrow A = PDP^{-1}$$

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD\cancel{(P^{-1}P)}DP^{-1} = PD^2P^{-1}$$

Via induction:

$$A^n = PD^nP^{-1}$$

$\star$

Look at the RHS of the formula $A^n = PD^nP^{-1}$. We know P and can easily calculate P^{-1}. The calculation of D^n is also trivial because D is such a simple diagonal matrix! This means we can actually find a formula for A^n. Let's do it:

Recall from the previous lecture that A has eigenvalues $\lambda = 3, -2$ with the associated eigenspaces being given by

$$E_3 = \text{span}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\} \text{ and } E_{-2} = \text{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}. \qquad A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$$

Hence if $P = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ we have a diagonalisation $P^{-1}AP = D$ where $D = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$.

Observe now that:

$$D^2 = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}$$

$$D^3 = \begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 27 & 0 \\ 0 & -8 \end{pmatrix}$$

and hence

$$D^n = \begin{pmatrix} 3^n & 0 \\ 0 & (-2)^n \end{pmatrix}$$

Since D is diagonal, D^n is just D with its non zero entries powered up.

It is difficult to power up the matrix A. But D is diagonal! It can be easily powered up, just by powering up its individual diagonal elements.

Using $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \dfrac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ we also have

$$P^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} = \frac{1}{2-1}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

Hence $A^n = PD^nP^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 3^n & 0 \\ 0 & (-2)^n \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$

$$= \begin{pmatrix} 2(3^n) & (-2)^n \\ 3^n & (-2)^n \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 2(3)^n - (-2)^n & -2(3)^n + 2(-2)^n \\ 3^n - (-2)^n & 2(-2)^n - 3^n \end{pmatrix} = A^n$$

Check:

$$A^0 = \begin{pmatrix} 2(3)^0 - (-2)^0 & -2(3)^0 + 2(-2)^0 \\ 3^0 - (-2)^0 & 2(-2)^0 - 3^0 \end{pmatrix} = \begin{pmatrix} 2-1 & -2+2 \\ 1-1 & 2-1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

$$A^1 = \begin{pmatrix} 2(3)^1 - (-2)^1 & -2(3)^1 + 2(-2)^1 \\ 3^1 - (-2)^1 & 2(-2)^1 - 3^1 \end{pmatrix} = \begin{pmatrix} 6+2 & -6-4 \\ 3+2 & -4-3 \end{pmatrix} = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix} = A.$$

The beauty of course, is that we can now power A up effortlessly. So A^{100} is quite simply:

$$A^{100} = \begin{pmatrix} 2(3^{100}) - (-2)^{100} & (-2)(3^{100}) + 2(-2)^{100} \\ 3^{100} - (-2)^{100} & 2(-2)^{100} - 3^{100} \end{pmatrix}$$

$$= \begin{pmatrix} 1030755041464022660805271659303013143907511838626 & -1030755041464022659537621059074783742410808633250 \\ 515377520732011329768810529537391871205404316625 & -515377520732011328501159929309162469708701111249 \end{pmatrix}$$

One simple calculation instead of 100 matrix multiplications!

★

The above method can be used to calculate A^n for any sized diagonalisable matrix A.

Homework Investigation: Does the above formula work when $n = -1$?

Using $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \dfrac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ we have

$$\begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}^{-1} = \frac{1}{-6}\begin{pmatrix} -7 & 10 \\ -5 & 8 \end{pmatrix} = \begin{pmatrix} \dfrac{7}{6} & -\dfrac{5}{3} \\[2ex] \dfrac{5}{6} & -\dfrac{4}{3} \end{pmatrix}.$$

Using $A^n = \begin{pmatrix} 2(3)^n - (-2)^n & -2(3)^n + 2(-2)^n \\ 3^n - (-2)^n & 2(-2)^n - 3^n \end{pmatrix}$ we have

$$A^{-1} = \begin{pmatrix} 2(3)^{-1} - (-2)^{-1} & -2(3)^{-1} + 2(-2)^{-1} \\ 3^{-1} - (-2)^{-1} & 2(-2)^{-1} - 3^{-1} \end{pmatrix} = \begin{pmatrix} \dfrac{2}{3} + \dfrac{1}{2} & -\dfrac{2}{3} - 1 \\[2ex] \dfrac{1}{3} + \dfrac{1}{2} & -1 - \dfrac{1}{3} \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{7}{6} & -\dfrac{5}{3} \\[2ex] \dfrac{5}{6} & -\dfrac{4}{3} \end{pmatrix} \qquad \text{....................amazing!}$$

★

So we can even use negative integers in the formula for A^n, provided of course that the matrix A has an inverse.

Example 2: Let

$$C = \begin{pmatrix} 3 & -1 \\ 5 & -\frac{3}{2} \end{pmatrix}$$

a) Find a formula for C^n.

b) Check your formula in a) at $n = 0$ and $n = 1$.

c) What is $\lim_{n \longrightarrow \infty} C^n$?

d) Suppose that $\mathbf{v} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ is **any** vector in $\mathbb{R}^2$.

Prove that $C^n \mathbf{v}$ converges to an eigenvector of C as n approaches infinity.

---

a) Be careful here. The examiner may not alert you to the fact that this is in fact an eigenvalue/eigenvector problem. You need to know the story! Our first job is to produce a diagonalisation of C via a complete eigenanalysis of the matrix.

a) $\begin{vmatrix} 3-\lambda & -1 \\ 5 & -3/2-\lambda \end{vmatrix} = 0$ $\qquad \det(C-\lambda I)=0$

$(3-\lambda)(-\frac{3}{2}-\lambda)+5 = 0$

$-\frac{9}{2} - 3\lambda + \frac{3}{2}\lambda + \lambda^2 + 5 = 0$

$\cancel{-9 - 6\lambda + 3\lambda + 2\lambda^2 + 10 = 0}$

$2\lambda^2 - 3\lambda + 1 = 0 \Rightarrow (2\lambda-1)(\lambda-1)=0$

$\therefore \lambda = 1, \frac{1}{2}$

check $\quad \Sigma \text{e-vals} = Tr(C)$

$1+\frac{1}{2} = 3 - 3/2 = 1\frac{1}{2}$ ✓✓

$\underline{\underline{\lambda = 1}}$: $\left(\begin{array}{cc|c} 2 & -1 & 0 \\ 5 & -5/2 & 0 \end{array}\right) \xrightarrow{R_2 = R_2 - 5/2 R_1} \left(\begin{array}{cc|c} 2 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right)$ ✓ x

Let $y = t$

$2x - t = 0$

$x = \frac{1}{2}t$

$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 1 \end{pmatrix} t$

In particular $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$

check $\begin{pmatrix} 3 & -1 \\ 5 & -3/2 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ ✓✓

$\therefore C\underset{\sim}{v} = 1\,\underset{\sim}{v}.$

$\underline{\lambda = \tfrac{1}{2}}$ $\quad\therefore\quad \begin{pmatrix} \frac{5}{2} & -1 \\ 5 & -2 \end{pmatrix}$ $\quad 3 - \frac{1}{2} \quad -3/2 - \frac{1}{2}$

$$\left(\begin{array}{cc|c} \frac{5}{2} & -1 & 0 \\ 5 & -2 & 0 \end{array}\right) \xRightarrow[R_2 = R_2 - 2R_1]{} \left(\begin{array}{cc|c} 5/2 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Let $y = t$

$5/2\,x - t = 0 \Rightarrow 5/2\,x = t \Rightarrow x = 2/5\,t$

$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2/5 \\ 1 \end{pmatrix} t$ and in particular

$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$ ✓✓

check $\begin{pmatrix} 3 & -1 \\ 5 & -3/2 \end{pmatrix}\begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 5/2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 \\ 5 \end{pmatrix}$

$10 - 15/2$

$\therefore C(\underset{\sim}{v}) = \frac{1}{2}\underset{\sim}{v}.$

$\therefore$ $P = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \Rightarrow P^{-1}CP = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$

$= D$

$$P^{-1}CP = D \Rightarrow C = PDP^{-1}$$
$$\Rightarrow C^n = PD^nP^{-1}$$

$$P = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \Rightarrow P^{-1} = \frac{1}{5 \cdot 4} \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}$$

$$C^n = P D^n P^{-1}$$

$$= \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1^n & 0 \\ 0 & (\frac{1}{2})^n \end{pmatrix} \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}$$

Now $\begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2}^n \end{pmatrix} = \begin{pmatrix} 1 & 2(\frac{1}{2})^n \\ 2 & 5(\frac{1}{2})^n \end{pmatrix}$

$$\begin{pmatrix} 1 & 2(\frac{1}{2})^n \\ 2 & 5(\frac{1}{2})^n \end{pmatrix} \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}$$

$$C^n = \begin{pmatrix} 5 - 4(\frac{1}{2})^n & -2 + 2(\frac{1}{2})^n \\ 10 - 10(\frac{1}{2})^n & -4 + 5(\frac{1}{2})^n \end{pmatrix}$$

b) $\underline{n=0}$: $\begin{pmatrix} 5-4 & -2+2 \\ 10-10 & -4+5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$n=1$: $\begin{pmatrix} 5-2 & -2+1 \\ 10-5 & -4+5/2 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ 5 & -3/2 \end{pmatrix} = C$ ✓✓

$$\star \quad P = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \Longrightarrow P^{-1}CP = D = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \quad \star$$

$$\star \quad a) \quad C^n = \begin{pmatrix} 5 - 4\left(\frac{1}{2}\right)^n & -2 + 2\left(\frac{1}{2}\right)^n \\ 10 - 10\left(\frac{1}{2}\right)^n & -4 + 5\left(\frac{1}{2}\right)^n \end{pmatrix} \quad c) \quad \begin{pmatrix} 5 & -2 \\ 10 & -4 \end{pmatrix} \quad \star$$

c) $\lim_{n\to\infty} C^n = \begin{pmatrix} 5 & -2 \\ 10 & -4 \end{pmatrix}$

d) $\begin{pmatrix} 5 & -2 \\ 10 & -4 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$$\lim_{n\to\infty} C^n \underset{\sim}{v} = \begin{pmatrix} 5 & -2 \\ 10 & -4 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$= \begin{pmatrix} 5\alpha - 2\beta \\ 10\alpha - 4\beta \end{pmatrix}$$

$$= (5\alpha - 2\beta)\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{e-vec}$$

which <u>is</u> an eigenvector!!

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 15

SYSTEMS OF DEs

Milan Pahor

MATH1231 ALGEBRA

SYSTEMS OF DEs

Suppose that A is a 2×2 matrix with two linearly independent eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ and associated eigenvalues λ_1, λ_2. Then the solution to the system of differential equations

$$\mathbf{y}' = A\mathbf{y}$$

takes the form

$$\mathbf{y} = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t}$$

where the two arbitrary constants c_1 and c_2 may be determined by applying initial conditions.

You have already examined the theory of differential equations in the calculus strand of Math1231. Remarkably it is also possible to use the theory of eigenvalues and eigenvectors to solve very complicated systems of differential equations.

It is not often that the worlds of algebra and calculus collide!

As with matrix powers in the previous lecture, we will base our exposition on the use of the familiar matrix

$$A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}.$$

Recall from the previous lecture that A has eigenvalues $\lambda = 3, -2$ with the associated eigenspaces being given by

$$E_3 = \text{span}\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\} \text{ and } E_{-2} = \text{span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}.$$

Systems of Differential Equations

Consider the system of differential equations:

$$\begin{aligned} {y_1}' &= 8y_1 - 10y_2 \qquad (1) \\ {y_2}' &= 5y_1 - 7y_2 \qquad (2) \end{aligned}$$

with initial conditions $y_1(0) = -1$ and $y_2(0) = 4$.

These systems are difficult to solve directly! We effectively have two intertwined first order differential equations. The solution takes the form of two functions y_1 and y_2 which satisfy both the initial conditions and the two differential equations.

But we can't solve equation (1) for y_1 since we don't know what y_2 is, and we can't use (2) to find y_2 since we don't know what y_1 is. It feels as if we are doomed to go around in circles forever, but we can crack the system using eigenvectors.

But where is the matrix?

Note that the system can be completely re-expressed in matrix-vector form as

$$\begin{pmatrix} {y_1}' \\ {y_2}' \end{pmatrix} = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

That is

$$\mathbf{y}' = A\mathbf{y}$$

where $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$ and $\mathbf{y}(0) = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$.

Amazingly, this system is trivial to solve once we have our hands on the eigenvectors and eigenvalues of the matrix A. The method of solution stems from a simple fact which we state on the next page and prove later. Note that, as usual, if the matrix A is short on eigenvectors, then the whole process fall apart and something else will need to be done.

Many of our second year courses analyse this pathological situation.

Note that, for simplicity, all of our examples in this lecture will be 2×2. The method of solution does however trivially extend to any diagonalisable $n \times n$ matrix.

Systems of DE's usually deal with structures evolving over time, so the independent variable is usually t. This is of no great concern.....you could use x instead.

Let's take a very careful look at a particular example.

Example 1: Solve the system of differential equations

$$\begin{aligned} {y_1}' &= 8y_1 - 10y_2 \qquad (1) \\ {y_2}' &= 5y_1 - 7y_2 \qquad (2) \end{aligned}$$

with initial conditions $y_1(0) = -1$ and $y_2(0) = 4$.

We begin by noting that everything can be rewritten in matrix vector form as:

$$\mathbf{y}' = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix} \mathbf{y}.$$

That is $\quad \mathbf{y}' = A\mathbf{y} \quad$ where $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, $A = \begin{pmatrix} 8 & -10 \\ 5 & -7 \end{pmatrix}$ and $\mathbf{y}(0) = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$.

Furthermore we already know from a previous lecture that the matrix A has eigenvalues $\lambda = 3, -2$ with the associated eigenspaces being given by

$$E_3 = \text{span}\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\} \text{ and } E_{-2} = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

That is it. We can now literally write down a solution to the system of DE's using the following claim, which we prove after the example:

> **Claim:** Suppose that A is a 2×2 matrix with two linearly independent eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ and associated eigenvalues λ_1, λ_2. Then the solution to the system of differential equations
>
> $$\mathbf{y}' = A\mathbf{y}$$
>
> takes the form
>
> $$\mathbf{y} = c_1 \mathbf{v}_1 e^{\lambda_1 t} + c_2 \mathbf{v}_2 e^{\lambda_2 t}$$
>
> where the two arbitrary constants c_1 and c_2 may be determined by applying initial conditions.

It hence follows that the solution takes the form

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = c_1 \mathbf{v}_1 e^{\lambda_1 t} + c_2 \mathbf{v}_2 e^{\lambda_2 t} = c_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} e^{3t} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-2t} = \begin{pmatrix} 2c_1 e^{3t} + c_2 e^{-2t} \\ \\ c_1 e^{3t} + c_2 e^{-2t} \end{pmatrix}.$$

Reading off the layers we have

$$\begin{aligned} y_1 &= 2c_1 e^{3t} + c_2 e^{-2t} \\ y_2 &= c_1 e^{3t} + c_2 e^{-2t} \end{aligned}$$

.... too easy!

Applying the initial conditions $y_1(0) = -1$ and $y_2(0) = 4$ yields

$$\begin{aligned} 2c_1 + c_2 &= -1 \\ c_1 + c_2 &= 4 \end{aligned}$$

with solution $c_1 = -5$ and $c_2 = 9$.

So our final solution to the systems of differential equations is

$$\begin{aligned} y_1 &= -10e^{3t} + 9e^{-2t} \\ y_2 &= -5e^{3t} + 9e^{-2t} \end{aligned}$$

★

We have found the two functions!

check: Let's check that the first differential equation $y_1' = 8y_1 - 10y_2$ is satisfied:

$LHS = y_1' = -30e^{3t} - 18e^{-2t}$.

$RHS = 8y_1 - 10y_2 = 8(-10e^{3t} + 9e^{-2t}) - 10(-5e^{3t} + 9e^{-2t})$

$= -80e^{3t} + 72e^{-2t} + 50e^{3t} - 90e^{-2t} = -30e^{3t} - 18e^{-2t} = y_1' = LHS.$

Hence the first differential equation is satisfied. The second may be checked similarly. Note also that $y_1(0) = -1$ and $y_2(0) = 4$ as required.

Before moving on to another example, let's prove that this all works!

Claim: Suppose that A is a 2×2 matrix with two linearly independent eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ and associated eigenvalues λ_1, λ_2. Then the solution to the system of differential equations

$$\mathbf{y}' = A\mathbf{y}$$

takes the form

$$\mathbf{y} = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t}$$

where the two arbitrary constants c_1 and c_2 may be determined by applying initial conditions.

Proof: Assume a solution to the system of DE's

$$\mathbf{y}' = A\mathbf{y}$$

of the form $\mathbf{y} = \mathbf{w}e^{kt}$ where $\mathbf{w}$ is a vector and k is a real number.

Then

$$\underset{\sim}{y} = \underset{\sim}{w}\, e^{kt}$$

$$LHS = \underset{\sim}{y}' = \underset{\sim}{w}\, e^{kt} k$$

$$RHS = A\underset{\sim}{y} = A\underset{\sim}{w}\, e^{kt}$$

$$\therefore LHS = RHS \implies \underset{\sim}{w}\, k\cancel{e^{kt}} = A\underset{\sim}{w}\,\cancel{e^{kt}}$$

$$\therefore A\underset{\sim}{w} = k\underset{\sim}{w}$$

$\therefore \underset{\sim}{w}$ is an eigenvector with eigenvalue k

Taking linear combinations of such solutions gives a general solution of the form

$$\mathbf{y} = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t}.$$

★

Note that the term $c_1\mathbf{v}_1 e^{\lambda_1 t}$ must have an eigenvalue λ_1 matched with its individual eigenvector $\mathbf{v}_1$. You can't mix and match the eigenvalues and eigenvectors!

Let's do another one from scratch.

Example 2: Determine the general solution to the system of differential equations

$$y_1{}' = 7y_1 + 4y_2 \quad (1)$$
$$y_2{}' = 4y_1 + y_2 \quad (2)$$

Check that your solutions for y_1 and y_2 satisfy differential equation (1).

Let $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

$$\therefore \quad \underset{\sim}{y}' = \begin{pmatrix} 7 & 4 \\ 4 & 1 \end{pmatrix} \underset{\sim}{y}$$

$$\therefore \quad A = \begin{pmatrix} 7 & 4 \\ 4 & 1 \end{pmatrix}$$

$$\begin{vmatrix} 7-\lambda & 4 \\ 4 & 1-\lambda \end{vmatrix} \Rightarrow (7-\lambda)(1-\lambda) - 16 = 0$$

$$\Rightarrow 7 - 7\lambda - \lambda + \lambda^2 - 16 = 0$$

$$\Rightarrow \lambda^2 - 8\lambda - 9 = 0$$

$$\Rightarrow (\lambda - 9)(\lambda + 1) = 0$$

$$\therefore \quad \lambda = -1, 9.$$

check Σ e-vals $= Tr(A)$

$\checkmark$ $-1+9 = 7+1$ $\checkmark$

$A\underset{\sim}{v} = \lambda \underset{\sim}{v}.$
$\Rightarrow (A - \lambda I)\underset{\sim}{v} = \underset{\sim}{0}$

$\underline{\underline{\lambda = -1}}$: $\left(\begin{array}{cc|c} 8 & 4 & 0 \\ 4 & 2 & 0 \end{array}\right)$

$R_2 = R_2 - \frac{1}{2}R_1 \quad \left(\begin{array}{cc|c} 8 & 4 & 0 \\ 0 & 0 & 0 \end{array}\right)$

Let $y = t$

$$8x + 4t = 0 \Rightarrow 8x = -4t$$
$$\Rightarrow x = -\tfrac{1}{2}t$$

$$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ 1 \end{pmatrix} t$$

In particular $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$

check $\begin{pmatrix} 7 & 4 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$

$$= -1 \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

$$\therefore A \begin{pmatrix} -1 \\ 2 \end{pmatrix} = (-1) \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

$\underline{\lambda = 9}$: $\left(\begin{array}{cc|c} -2 & 4 & 0 \\ 4 & -8 & 0 \end{array}\right)$

$R_2 = R_2 + 2R_1$ $\begin{array}{c} \begin{array}{cc} \checkmark & \times \end{array} \\ \left(\begin{array}{cc|c} -2 & 4 & 0 \\ 0 & 0 & 0 \end{array}\right) \end{array}$

Let $y = t$

$$-2x + 4t = 0 \Rightarrow x = 2t$$

$$\therefore \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} t$$

In particular $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$$\begin{pmatrix} 7 & 4 \\ 4 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 18 \\ 9 \end{pmatrix}$$

$$= 9\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$\therefore \quad A\begin{pmatrix} 2 \\ 1 \end{pmatrix} = 9\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$\therefore \quad E_9 = sp\left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right\}, \quad E_{-1} = sp\left\{\begin{pmatrix} -1 \\ 2 \end{pmatrix}\right\}$$

$$\therefore \quad \underset{\sim}{y} = c_1\begin{pmatrix} 2 \\ 1 \end{pmatrix}e^{9t} + c_2\begin{pmatrix} -1 \\ 2 \end{pmatrix}e^{-t} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$y_1 = 2c_1 e^{9t} - c_2 e^{-t}$$

$$y_2 = c_1 e^{9t} + 2c_2 e^{-t}$$

---

check $y_1' = 7y_1 + 4y_2$

$$LHS = y_1' = 18c_1 e^{9t} + c_2 e^{-t}$$

$$RHS = 7y_1 + 4y_2 = 7\{2c_1 e^{9t} - c_2 e^{-t}\} + 4\{c_1 e^{9t} + 2c_2 e^{-t}\}$$

$$= 14c_1 e^{9t} - 7c_2 e^{-t} + 4c_1 e^{9t} + 8c_2 e^{-t}$$

$$= 18c_1 e^{9t} + c_2 e^{-t}$$

$$= LHS$$

$$* \quad \begin{aligned} y_1 &= 2c_1e^{9t} - c_2e^{-t} \\ y_2 &= c_1e^{9t} + 2c_2e^{-t} \end{aligned} \quad *$$

That is the end of the algebra! The next lecture will be a revision of set theory and elementary probability theory (2 unit). The rest of the course will then explore various advanced topics in probability and statistics.

Homework: Connecting Math1231 Calculus with Math1231 Algebra

To finish off, let's solve a simple second order differential equation in two very very different ways.

Example 3: Solve the constant coefficient second order differential equation

$$y'' - 5y' + 6y = 0.$$

Math1231 Calculus Method: Aux eqn $\lambda^2 - 5\lambda + 6 = 0$

$(\lambda - 2)(\lambda - 3) = 0 \Rightarrow \lambda = 2, 3$

$\therefore y = c_1 e^{2t} + c_2 e^{3t}$

$$\star \quad y = c_1 e^{2t} + c_2 e^{3t} \quad \star$$

Math1231 Algebra Method: Interestingly, the single D.E. $y'' - 5y' + 6y = 0$ is not a system of equations, but can be converted over into a system!

Let $u_1 = y$ and $u_2 = y'$. Then we have

${u_1}' = u_2.$

Also

${u_2}' = y'' = -6y + 5y' = -6u_1 + 5u_2.$

We therefore have the system of differential equations

$$\begin{aligned} {u_1}' &= 0u_1 + 1u_2 \\ {u_2}' &= -6u_1 + 5u_2 \end{aligned}$$

which can be rewritten in matrix-vector form as:

$$\mathbf{u}' = \begin{pmatrix} 0 & 1 \\ -6 & 5 \end{pmatrix} \mathbf{u}.$$ where $\underset{\sim}{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$

That is

$$\mathbf{u}' = A\mathbf{u}$$

where $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ and $A = \begin{pmatrix} 0 & 1 \\ -6 & 5 \end{pmatrix}$.

It is easily shown that the matrix A has

eigenvalues $\lambda = 2, 3$ with the associated eigenvectors

$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 3\ . \end{pmatrix}$.

Thus the solution takes the form

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = c_1 \mathbf{v}_1 e^{\lambda_1 t} + c_2 \mathbf{v}_2 e^{\lambda_2 t} = c_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} e^{2t} + c_2 \begin{pmatrix} 1 \\ 3 \end{pmatrix} e^{3t}.$$

Hence reading off the layers we have

$$\begin{aligned} u_1 &= c_1 e^{2t} + c_2 e^{3t} \\ u_2 &= 2c_1 e^{2t} + 3c_2 e^{3t} \end{aligned}$$

But $u_1 = y$. So

$y = c_1 e^{2t} + c_2 e^{3t}$, same as before.

The u_2 is nothing but y' and can be tossed in the bin.

Obviously the calculus methods are superior for solving constant coefficient second order differential equations. But when it comes to **systems** of DE's, eigenvectors are the way to go!

★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 16

REVISION OF BASIC PROBABILITY

Milan Pahor

MATH1231 ALGEBRA

REVISION OF BASIC PROBABILITY

The set of all possible outcomes of a given experiment is called the **Sample Space** for that experiment and will be denoted by S.

An **Event** E is simply **a subset of the sample space.** That is $E \subseteq S$.

A simple definition of the probability $P(E)$, of an event E is

$$P(E) = \frac{|E|}{|S|} = \frac{\text{number of elements in } E}{\text{number of elements in } S}.$$

See your printed notes for a more formal definition.

Rules for Probability:

1) $0 \leq P(E) \leq 1$.

2) If $P(E) = 1$ then E **must** happen. If $P(E) = 0$ then E **cannot** happen.

3) $P(E^c) = 1 - P(E)$.

That is, the probability of something not happening is one minus the probability of it happening.

4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

In this lecture we will revise some relevant high school set and probability theory, in preparation for our future analysis of discrete and continuous random variables.

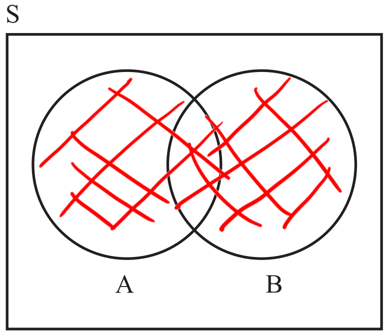
The concept of a **set** sits at the heart of probability theory.

Definition: A **set** is an unordered collection of objects. We call the individual objects **elements** of the set. We **always** use curly brackets {........} to describe sets.

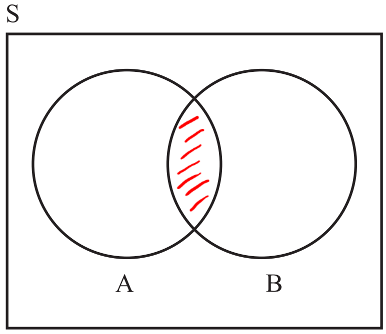
Set Operations and Concepts:

Let A and B be sets. We define the following operations and concepts, displaying the outcome on a Venn diagram:

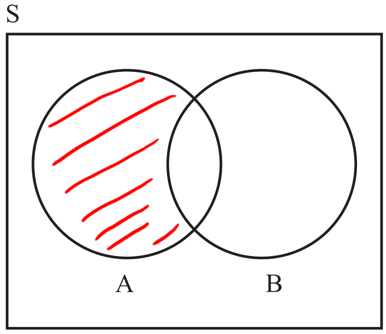
a) The **union** of two sets A and B, denoted by $A \cup B$, is the set that contains those elements that are either in A or in B or both.



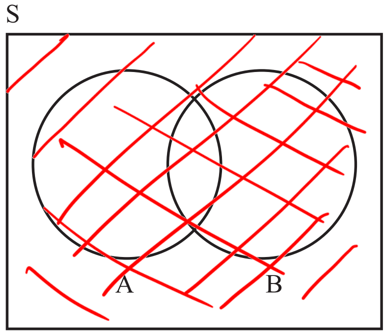
b) The **intersection** of two sets A and B, denoted by $A \cap B$, is the set that contains those elements that are both in A and in B.



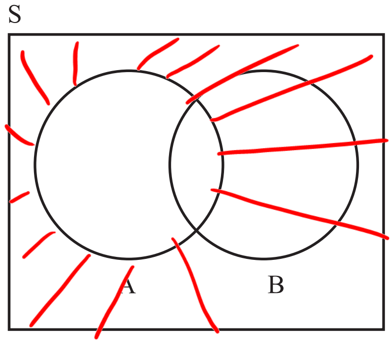
c) The **difference** of A and B, denoted by $A - B$ is the set that contains those elements that are in A but not in B.



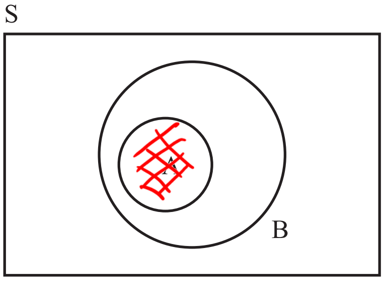
d) The **universal set** S is a set containing all elements of all sets that we will need to talk about in some particular topic.



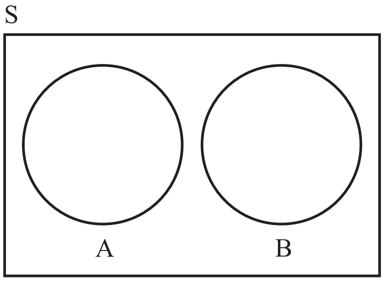
e) The **complement** of the set A, denoted by A^c, is $S - A$.



f) The set A is said to be a **subset** of B if and only if every element of A is also an element of B. We use the notation $A \subseteq B$ to indicate that A is a subset of B and $A \subset B$ to denote that A is a subset of B, but not equal to B.



g) Two sets are called **disjoint** if their intersection is empty, that is, A and B are disjoint if $A \cap B$ is the empty set ϕ.



We always denote the empty set, that is, the set with no elements, by ϕ. Strangely, the empty set is one of the most important sets in all of set theory!

Example 1: Let the universal set S be $\{a, b, c, d, e, f, g, h\}$ with subsets $A = \{a, c, e, h\}$, $B = \{c, d, f\}$ and $C = \{a, b, c, d, f, g\}$.

Write down:

i) $A \cup B$. $= \{a, c, e, h, d, f\}$

ii) $A \cap B$. Are A and B disjoint? $= \{c\} = A \cap B$

iii) $A - B$. $= \{a, e, h\}$

iv) A^c. $= \{b, f, g\}$

v) Is $B \subseteq C$? Yes

vi) Is $B \subseteq B$? Yes

vii) Is $B \subset B$? No

★

We write $|A|$ for the number of elements in a finite set A. Some sources will write $n(A)$.

Example 2: If $A = \{$all the letters in the English alphabet $\}$, what is $|A|$?

$|A| = 26$

★

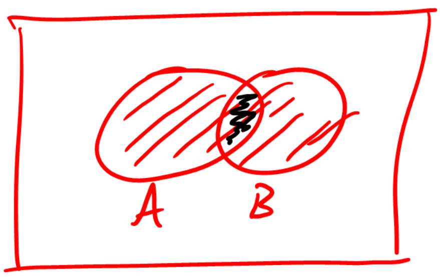
Theorem: If A and B are finite sets then

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Proof:

Consider the following Venn diagram.

$|A \cup B|$
$= |A| + |B| - |A \cap B|$



When we combine the elements of A and B to produce $|A| + |B|$, those elements in the intersection $A \cap B$ are added twice. Thus we need to take away $|A \cap B|$ once. ★

We generally use Venn diagrams to solve problems involving the size of $A \cup B$.

Probability

We will use the word *experiment* to describe the recording of an observation or a measurement of something that happens. For example, tossing a die and then recording the outcome is an experiment.

Definition:

• The set of all possible outcomes of a given experiment is called the **Sample Space** for that experiment and will be denoted by S.

• An **Event** E is simply **a subset of the sample space.** That is $E \subseteq S$.

A simple definition of the probability $P(E)$ of an event E is then

$$P(E) = \frac{|E|}{|S|} = \frac{n(E)}{n(S)} = \frac{\text{number of elements in } E}{\text{number of elements in } S}.$$

See your printed notes for a more formal definition.

Rules for Probability:

1) $0 \leq P(E) \leq 1$.

2) If $P(E) = 1$ then E **must** happen. If $P(E) = 0$ then E **cannot** happen.

3) $P(E^c) = 1 - P(E)$.

That is, the probability of something not happening is one minus the probability of it happening.

4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Example 3: A number is randomly chosen from the first ten positive integers.

a) Describe the sample space S. $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

b) Find the probability of selecting a number which is a multiple of 3.

$3/10 = \frac{n(E)}{n(S)}$

c) Find the probability of selecting a number which is a not a multiple of 3.

$1 - 3/10 = 7/10$

d) Find the probability of selecting a number which is a multiple of 3 or 4.

$5/10 = \frac{1}{2}$

e) Find the probability of selecting a number which is a multiple of 3 and 4.

$\frac{0}{10} = 0.$

Generally the word "and" leads to a multiplication of probabilities, while the word "or" leads to addition of probabilities. Generally this depends upon the intersection and independence of the events, which will be dealt with in the next lecture.

Example 4: Bob and Jane shoot once each at a target. The probability that Bob hits the target is 0.3 while the probability that Jane hits the target is 0.6. Find the probability that:

a) They both hit the target.

b) Bob hits and Jane misses.

c) Exactly one of them hits the target. (Do this in two different ways)

a) $P(B \underline{\text{and}}\ J) = 0.3 \times 0.6 = 0.18$

b) $P(B \text{ and } J^c) = 0.3 \times 0.4 = 0.12$

c) $P(B \text{ and } J^c \text{ OR } B^c \text{ and } J) = .3 \times .4 + .7 \times .6 = 0.54$

Method: $1 - P(B \text{ hits and } J \text{ hits OR } B \text{ misses and } J \text{ misses})$

$= 1 - \{.3 \times .6 + .7 \times .4\} = 1 - .18 - .28 = 0.54$

★

Example 5: Bob from the previous question takes 8 shots at the target. What is the probability he gets at least one hit?

$P(\text{at least one hit}) = 1 - P(\text{no hits})$

$= 1 - (.7)^8 = \underline{0.942}$

★ $1 - (.7)^8 = 0.942$ ★

Example 6: Bob from the previous question takes n shots at the target. How many shots does he need to take in order for the probability of at least one hit to be at least 0.999?

$P(\text{at least 1 hit}) = 1 - P(\text{no hits})$

$= 1 - (.7)^n \geq 0.999$

$\Rightarrow (.7)^n \leq 0.001 \Rightarrow \ln((.7)^n) \leq \ln(0.001)$

$\Rightarrow n \ln(0.7) \leq \ln(0.001) \Rightarrow n \geq \dfrac{\ln(0.001)}{\ln(0.7)}$

★ $n \geq \dfrac{\ln(0.001)}{\ln(0.7)} \Longrightarrow n \geq 19.367 \Longrightarrow n \geq 20$ ★

There are three distinct types of basic probability questions we need to be able to solve.

(I) Venn Diagram Questions: These involve overlapping groups.

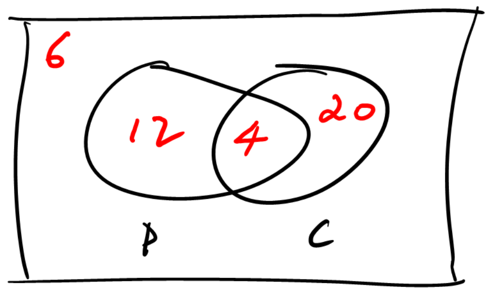
(II) Sample Space Questions: Often tossing two dice.

(III) Tree diagram questions: Usually selecting balls in boxes.

(I) **Example 7**: In a group of 42 students, 16 study Physics, 24 study Chemistry and 6 study neither.

a) How many study both.

b) What is the probability that a randomly selected student studies Physics only?



a) 4

b) $\frac{12}{42} = \frac{6}{21} = \frac{2}{7}$

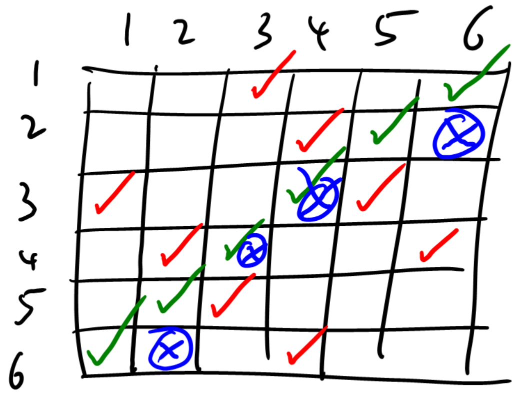
★ $a)$ 4 $b)$ $\frac{2}{7}$ ★

(II) **Example 8**: Two standard six sided dice are tossed and the number uppermost on each is noted. Find the probability that the two numbers:

a) Differ by 2.

b) Sum to 7.

c) Have a product of 12.



a) $\frac{8}{36} = \frac{4}{18} = \frac{2}{9}$

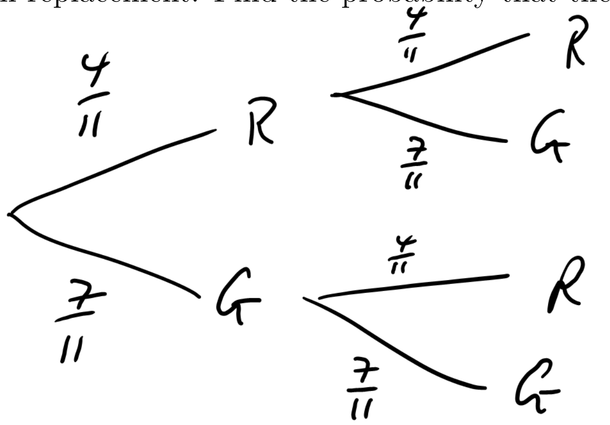
b) $\frac{6}{36} = \frac{1}{6}$

c) $\frac{4}{36} = \frac{1}{9}$

★ $a)$ $\frac{2}{9}$ $b)$ $\frac{1}{6}$ $c)$ $\frac{1}{9}$ ★

Observe that the events, when viewed on the grid, usually form a clearly defined pattern.

(III) **Example 9**: A box contains 4 red balls and 7 green balls. Two balls are selected with replacement. Find the probability that they are of the same colour.



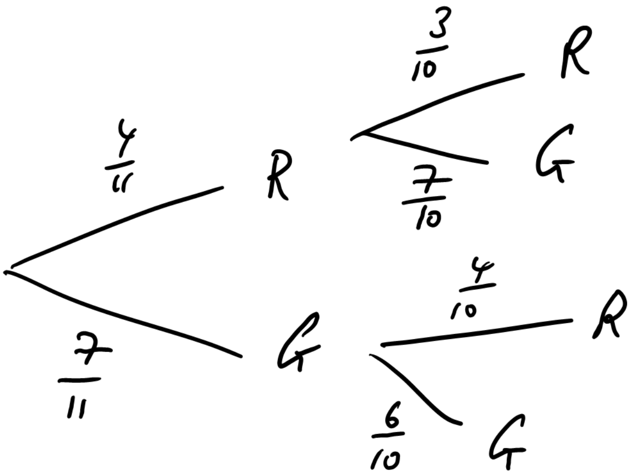
and $\rightarrow \times$
or $\rightarrow +$

P(R and R OR G and G)

$$\frac{4}{11}\times\frac{4}{11} + \frac{7}{11}\times\frac{7}{11} = \frac{16+49}{121} = \frac{65}{121}$$

$\star \quad \frac{65}{121} \quad \star$

Example 10: Repeat the above example with the selection being without replacement.



P(R and R OR G and G)

$$= \frac{4}{11}\times\frac{3}{10} + \frac{7}{11}\times\frac{6}{10} = \frac{12+42}{110} = \frac{54}{110} = \frac{27}{55}.$$

$\star \quad \frac{27}{55} \quad \star$

We will revisit these last examples in the next lecture from the point of view of conditional probability. The next lecture will also cover the use of permutations and combinations in probability.

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 17

CONDITIONAL PROBABILITY

Milan Pahor

MATH1231 ALGEBRA
CONDITIONAL PROBABILITY

If two events A and B are independent, we have

$$P(A \cap B) = P(A \text{ and } B) = P(A).P(B).$$

Conditional probabilities for dependent events are calculated via

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \text{ and } B)}{P(B)}.$$

Independence

We say that two events A and B are **independent** of each other if the probability that one of them occurs is not influenced by the occurrence or non-occurrence of the other. They have nothing to do with each other!

For example, if we roll a die, and then toss a coin, then the outcome of the coin toss is independent of the outcome of the rolling of the die.

If two events A and B are independent, we have

$$P(A \cap B) = P(A \text{ and } B) = P(A).P(B).$$

That is , for independent events we just multiply the probabilities out for sequential outcomes.

Example 1: A fair die is rolled, followed by a toss of an unbiased coin. Find the probability of getting a 5 on the die and a head on the coin.

$$\frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$$

★

But not all events are independent! For example the event

$A \equiv$ {I carry an umbrella to work} and the event

$B \equiv$ {It is raining when I leave the house}

are clearly dependent!

The probability of A occurring certainly depends upon the outcome of B. In these circumstances we talk in terms of the **conditional** probability of A occurring, given that B has already happened. Let's revisit some examples from the last lecture and ask some conditional probability questions.

We denote the conditional probability of event A given that B has occurred by

$$P(A|B)$$

which is read as "the probability of A given B".

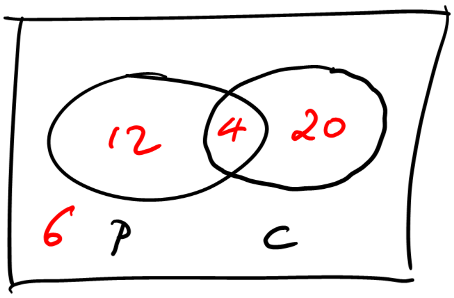
Example 2: In a group of 42 students, 16 study Physics, 24 study Chemistry and 6 study neither.

a) How many study both?

b) What is the probability that a student studies Physics?

c) What is the probability that a student studies Physics, given that they study Chemistry? That is, what is

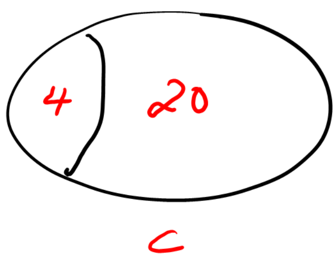
$$P(\text{ student studies Physics } | \text{ they study Chemistry })?$$



a) 4

b) $\frac{16}{42} = \frac{8}{21}$

c)



$\frac{4}{24}$

$= \frac{1}{6}$

★ *a)* 4 *b)* $\frac{8}{21}$ *c)* $\frac{1}{6}$ ★

Observe that, once we knew that the student studied chemistry for sure, our focus restricted from the entire diagram down to the chemistry bubble. That is our sample space was cut down from 42 elements to only 24 elements. This is how conditional probability works!

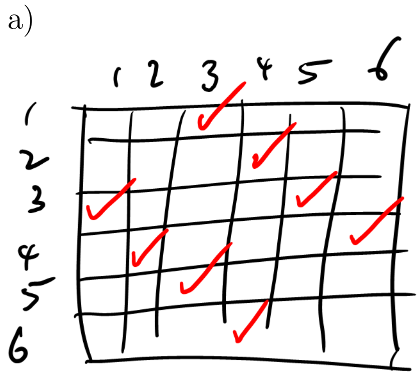
(II) **Example 3**: Two standard six sided dice are tossed and the number uppermost on each is noted. Find the probability that the two numbers:

a) Differ by 2.

b) Differ by 2, given that at least one of the numbers is 3. That is, calculate

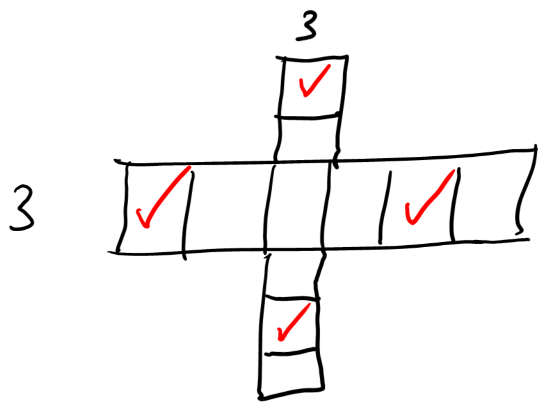
$$P(\text{ Numbers differ by 2} \mid \text{At least one of them is a 3 })?$$

a)



$$\frac{8}{36} = \frac{4}{18} = \frac{2}{9}$$

b)



$$P(\text{differ by 2} \mid \text{at least one is a 3}) = \frac{4}{11}$$

★ $a)\ \frac{2}{9}$ $b)\ \frac{4}{11}$ ★

Observe once again how our sample space has been restricted from a full set of 36 elements to a subset of only 11.

Let's take another look at the calculation of the conditional probability in the previous example.

$$P(\text{ Numbers differ by 2} \mid \text{One of them is a 3 }) = \frac{4}{11} = \frac{\frac{4}{36}}{\frac{11}{36}}$$

$$= \frac{\text{Probability that the numbers differ by two and one of them is a 3}}{\text{Probability one of them is a 3}}.$$

This suggests the following lovely little formula for conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \text{ and } B)}{P(B)}.$$

That is, to calculate the probability that A occurs given that B has happened, first restrict your sample space down to B (this is the $P(B)$ in the bottom) and then pick off the elements of your new sample space which are of interest. This is the $P(A \text{ and } B)$ on the top.

Example 4: Urn A contains 9 white balls and 1 red ball, while Urn B contains 2 white balls and 3 red balls. An urn is selected at random and a single ball is then chosen from the selected urn.

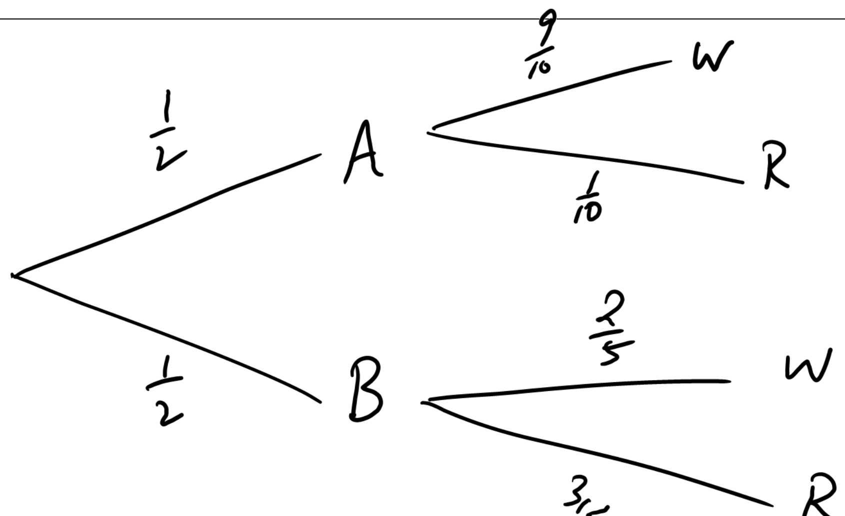
a) Draw a tree diagram to represent this experiment.

b) What is the probability of choosing a white ball?

c) If the chosen ball is white, what is the probability that it came from urn A?

---

a)



b)

$$P(\text{white}) = P(A \text{ and } W \text{ OR } B \text{ and } W)$$

$$= \frac{1}{2} \times \frac{9}{10} + \frac{1}{2} \times \frac{2}{5} = \frac{9}{20} + \frac{2}{10}$$

$$= \frac{13}{20}.$$

c) This is quite an interesting question.

A naive answer would be that, since the urns were selected with equal probability, the probability that the white ball came from Urn A is $\dfrac{1}{2}$.

But say Urn B had **no** white balls. Then the eventual selection of a white ball would imply that it came from Urn A with a probability of 1 not $\frac{1}{2}$, even though the urns were selected with equal probability! Clearly we need to be careful here! We have:

$$P(\text{Urn A} \mid \text{White ball selected}) = \frac{P(\text{Urn A and White ball selected})}{P(\text{White ball selected})}.$$

$$= \frac{\frac{1}{2} \times \frac{9}{10}}{\frac{1}{2} \times \frac{9}{10} + \frac{1}{2} \times \frac{2}{5}} = \frac{\frac{9}{20}}{\frac{13}{20}} = \frac{9}{13} \approx 0.69$$

$$\bigstar \qquad b)\ \frac{13}{20} \quad c)\ \frac{9}{13} \approx 0.69 \qquad \bigstar$$

There are a lot of points to note here!

• The probability that Urn A was the source of the white ball is more than 0.5 since most of the white balls are with Urn A. The fact that a white was chosen makes you lean towards Urn A as the source.

• Problems of this type, where conditional probability is extracted backwards through a tree diagram, are often referred to as Bayes' rule questions, and you will find a Bayes' rule formula in your printed notes. Rather than using a formula it is best however just to consider a tree diagram and use the general conditional probability equation.

• Observe that the calculations in the numerator are a subset of those in the denominator. This always happens with Bayes' rule questions.

Discussion: What becomes of the formula

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}. \qquad = \frac{P(A)\cancel{P(B)}}{\cancel{P(B)}}$$

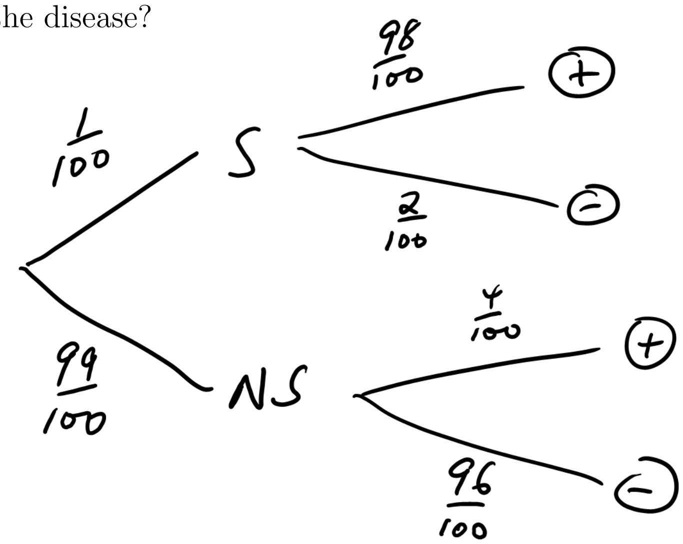
if A and B are independent events?

$$= P(A)$$

$\bigstar$

The issue of conditional probability is a critical tool in evaluating diagnostic medical tests.

Example 5: It is known that 1% of a given population suffer from *pretendanitis*. If an individual has *pretendanitis* a diagnostic test will detect the disease with a correct positive result in 98% of cases. (Note that a positive test result is bad news). If an individual does not have *pretendanitis* the same test will mistakenly detect the disease, giving a false positive, 4% of the time. Suppose that a randomly selected individual tests positive for *pretendanitis*. Use Bayes' theorem to determine the probability that they actually have the disease?



$$P(S/\oplus) = \frac{P(S \text{ and } \oplus)}{P(\oplus)}$$

$$= \frac{\frac{1}{100} \times \frac{98}{100}}{\frac{1}{100} \times \frac{98}{100} + \frac{99}{100} \cdot \frac{4}{100}}$$

$$= \frac{98}{98 + 4 \times 99} = 0.19838$$

★ .19838 ★

The test is useless! Only 20% of the people who test positive actually have the disease! In other words 80% of people who are told that they have *pretendanitis* are in fact healthy.

A false positive rate of 4% is problematic since the vast bulk of individuals are in fact healthy and the false positives are swamping the genuine cases. Tests for rare diseases need to be extremely accurate!

We finish off with a little revision on perms/comms and probability.

Permutations

For permutation questions **order** is crucial. You will see terms like "line-up" and "arrangement".

Example 6: Display the number of possible ways three people A, B and C can line up for a bus. How could you evaluate the total number of arrangements without a list?

ABC, ACB, BAC, BCA, CAB, CBA

$\therefore 6$ ★

Example 7: Count the number of possible ways thirty people can line up for a bus.

We just fill boxes!

$\boxed{30}\ \boxed{29}\ \boxed{28}\ \boxed{27}\ \boxed{26} \ldots\ldots \boxed{1}$

$30 \times 29 \times 28 \times 27 \times \ldots\ldots 1$ ★ $30!$ ★

Definition: $n! = (n)(n-1)(n-2)\ldots(1)$ for any integer $n \geq 0$.

Fact: $n!$, read as n factorial, is the number of ways of permuting n distinct objects.

Example 8: Evaluate each of the following, first by hand and then via the calculator.

a) $3! = 3 \times 2 \times 1 = 6$

b) $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

c) $0! = 1$

d) $100! =$ XX Too big for calculator.

★

Example 9: Simplify each of the following:

a) $\dfrac{7!}{6!}$.

b) $\dfrac{n!}{(n+2)!}$.

★

a) $\dfrac{7!}{6!} = \dfrac{7 \times \cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times \cancel{1}}{\cancel{6} \times \cancel{5} \times \cancel{4} \times \cancel{3} \times \cancel{2} \times 1} = 7.$

b) $\dfrac{n!}{(n+2)!} = \dfrac{1}{(n+2)(n+1)}$

$P(E) = \frac{n(E)}{n(S)}$

Example 10: Four adults and 5 children line up randomly for a photo. Calculate the probability that there are children on both ends.

Method 1: P(children both ends)

$\boxed{5}\boxed{7}\boxed{6}\boxed{5}\boxed{4}\boxed{3}\boxed{2}\boxed{1}\boxed{4} = \frac{5 \times 4 \times 7!}{9!} = \frac{5 \times 4}{9 \times 8} = \frac{5}{18}$

Method 2:

$\frac{5}{9} \cdot \frac{4}{8} = \frac{5}{18}$

$\bigstar \quad \frac{5}{18} \quad \bigstar$

Combinations

For combination questions there is no **order**. You will see terms like "team", "group" and "arrangement".

The key to evaluating combination questions is the use of the binomial coefficient ${}^nC_r$, read as n choose r. Evaluation of ${}^nC_r$ can be accomplished through its definition

$$ {}^nC_r = \frac{n!}{r!(n-r)!} $$

or simply via the calculator.

Note that ${}^nC_r$ is also written as $\binom{n}{r}$.

Example 11: Evaluate ${}^9C_4$ by hand and also by calculator.

${}^9C_4 = \frac{9!}{4!\,(9-4)!} = 126$

OR Calculator

$\bigstar \quad 126 \quad \bigstar$

Fact: The number of different teams of r from a group of n is given by ${}^nC_r$.

Example 12: How many teams of three can be selected from the five people $\{A, B, C, D, E\}$.

Note that we have no order here, there is no first person and no last person.....just teams.

Quick way: ${}^5C_3 = 10$

$$\frac{5!}{3!(5-3)!} = \frac{120}{6.2} = \frac{120}{12} = 10$$

★ $\{ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE\} = 10$ ★

Example 13: How many teams of two can be selected from the five people $\{A, B, C, D, E\}$. Explain your answer.

${}^5C_2 = 10$

${}^5C_3 = {}^5C_2$

${}^nC_r \equiv {}^nC_{n-r}$

★

Example 14: A golf team of 4 players is randomly selected from 5 Australians and 7 Bulgarians. Find the probability that it comprises:

3A and 1B OR 4A 0B

a) Only Australians.

$$\frac{{}^5C_4 \times {}^7C_0}{{}^{12}C_4} = \frac{1}{99}$$

b) More Australians than Bulgarians.

$$\frac{{}^5C_3 \times {}^7C_1 + {}^5C_4 \times {}^7C_0}{{}^{12}C_4}$$

$$= \frac{5}{33}$$

★ $a) \quad \frac{{}^5C_4 \times {}^7C_0}{{}^{12}C_4} = \frac{1}{99} \qquad b) \quad \frac{{}^5C_4 \times {}^7C_0 + {}^5C_3 \times {}^7C_1}{{}^{12}C_4} = \frac{5}{33}$ ★

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 18

DISCRETE RANDOM VARIABLES

Milan Pahor

MATH1231 ALGEBRA

DISCRETE RANDOM VARIABLES

The expected value (also called the mean) of a discrete random variable X is given by

$$\mu = E(X) = x_1p_1 + x_2p_2 + \ldots x_np_n = \sum x_kp_k$$

$$E(aX + b) = aE(X) + b$$

In many of our previous examples, the fundamental sample space probabilities were flat. For example, when tossing two dice, there are 36 different possible outcomes, all with identical probabilities of $\frac{1}{36}$ of occurring.

But in reality, probabilities are distributed in a much more uneven fashion. In this lecture we will look carefully at the probability distributions for various discrete random variables. This will lead us naturally to continuous random variables and then the Normal distribution in upcoming lectures.

Definition: A random variable (usually denoted X), is a variable whose value is determined by chance.

We usually denote random variables with capital letters X, Y, Z etc.

There are two types of random variables:

• Discrete random variables can only take on a countable number of distinct separate values. For example cars in a parking lot, hairs on a head, numbers on a die.

• Continuous random variables can take on any value within a range. For example height of a giraffe or length of a time interval.

We will spend the next two lectures looking at the discrete case and then move on to continuous random variables.

A crucial feature of discrete random variables is that the probability of different outcomes need not be the same. We keep track of the probabilities through what is called a discrete **probability distribution**, usually presented as a table.

The top row of the distribution table is all possible outcomes for the random variable X. The bottom row displays the respective probabilities $P(X = x_k)$ of their occurrence.

Note that $P(X = x_k)$ is also denoted by p_k.

Example 1: A discrete random variable X has the following probability distribution:

x_k	1	2	4	9
$P(X = x_k)$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{4}$

Calculate:

a) $P(X = 4)$. $= \frac{1}{3}$

b) $P(X = 5)$. 0

c) $P(X > 2)$. $\frac{1}{3} + \frac{1}{4} = \frac{7}{12}$

d) $P(X \geq 2)$. $\frac{10}{12} = 5/6$

e) The sum of all the probabilities. $\sum p_k = 1$

★

Note that any probability distribution **MUST** have two critical properties:

1) The probabilities must be non-negative.

2) The total sum of all the probabilities must be equal to 1. That is, $\sum p_k = 1$.

Example 2: Prove that the following probability distribution table is valid:

x_k	-1	-2	3
$P(X = x_k)$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{15}{27}$

i) $p_k \geq 0$

(ii) $\sum p_k = \frac{3+9+15}{27} = \frac{27}{27} = 1$ ✓

★

Observe that the outcomes most certainly are allowed to be negative! However the probabilities of course are not.

Example 3: The total number of cavities X a patient has during their first visit to a particular dentist has the following probability distribution table:

x_k	0	1	2	3	4
p_k	5α	10α	3α	α	α

a) Find α.

b) What is the probability that a new patient has no cavities?

a) $\sum p_k = 20\alpha = 1 \Rightarrow \alpha = \frac{1}{20}$

b) $5\alpha = \frac{5}{20} = \underline{\underline{\frac{1}{4}}}$

★

Example 4: On her income tax return, Sally claimed her business trip expenses X were described by the following probability distribution table:

x_k	\$3000	\$1000	\$5000	\$10000	\$15000
$P(X = x_k)$	$\alpha - 2$	6α	8α	7α	8α

Prove that Sally is lying.

$\sum p_k = \alpha - 2 + 29\alpha = 30\alpha - 2 = 1$

$30\alpha = 3 \Rightarrow \alpha = \frac{1}{10}$

☹

$\alpha - 2 = \frac{1}{10} - 2 \underline{\underline{< 0}}$

★

Probability distributions for a random variable X are often formulated by using the techniques of the previous two lectures.

Example 5: Two fair standard dice are tossed and the random variable X has a value equal to the sum of the two numbers which appear on the uppermost faces.

a) Create a probability distribution table for X.

b) Check that $\sum p_k = 1$.

a)

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

x_k	2	3	4	5	6	7	8	9	10	11	12
$P(X=x_k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

b) $\Sigma p_k = \frac{36}{36} = 1$

★

x_k	2	3	4	5	6	7	8	9	10	11	12
$P(X=x_k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

★

Expected Value E(X)

When tossing a standard die the set of all possible outcomes is

$$D = \{1, 2, 3, 4, 5, 6\}.$$

and each of these six outcomes is of course equally likely.

We say that such a flat distribution of probabilities is **uniform** or that the associated discrete random variable X has a uniform distribution.

The average μ of the numbers in D is easy to find:

$$\mu = \frac{1+2+3+4+5+6}{6} = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5.$$

But what if the toss of the die is **not** fair? Say that it is governed by the probability distribution

x_k	1	2	3	4	5	6
p_k	0.01	0.01	0.01	0.01	0.95	0.01

Then the average surely is no longer 3.5! It should instead be very close to 5, since 5 dominates the possibility of making a contribution.

In this circumstance, we call the average the mean or the expected value and denote it by μ or $E(X)$.

The expected value $E(X)$ of a random variable X is a measure of the centre of the distribution. It is what you would *expect* to happen. When calculating the expected value $E(X)$ of a random variable X we need to take into account not only the outcomes, but also the probability of those outcomes making a contribution. The formula for $E(X)$ is simple and elegant:

Definition: The expected value (also called the mean) of a discrete random variable X is given by

$$\mu = E(X) = x_1 p_1 + x_2 p_2 + \ldots x_n p_n = \sum x_k p_k$$

In other words $E(X)$ is the total sum of all the outcomes multiplied (weighted) by their probability of occurrence.

Example 6: Find the expected value of the loaded die above.

$1 \times 0.01 + 2 \times .01 + 3 \times .01 + 4 \times .01 + 5 \times .95 + 6 \times .01$

★ 4.91 ★

Observe that the expected value of a random variable does **NOT** need to be one of the outcomes, and in fact $E(X)$ is rarely one of the x_k's in the first row of the distribution table.

Example 7: Find the expected value of the sum X of the two uppermost numbers when tossing two fair dice.

Recall from above that the probability distribution of X is:

x_k	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Thus the expected value is

$$E(X) = 2\times\frac{1}{36}+3\times\frac{2}{36}+4\times\frac{3}{36}+5\times\frac{4}{36}+6\times\frac{5}{36}+7\times\frac{6}{36}+8\times\frac{5}{36}+9\times\frac{4}{36}+10\times\frac{3}{36}+11\times\frac{2}{36}+12\times\frac{1}{36} = 7.$$

★

This means that when playing backgammon you will on average move 7 pips per roll.

Not surprisingly this is twice 3.5, which was the average outcome for a single die.

Example 8: Let X be the random variable in Example 7 and define a new random variable Y by $Y = 3X + 4$. Find $E(Y)$.

We have a simple formula for the expected value of linear transformations of random variables:

$$E(aX + b) = aE(X) + b$$

In other words the expectation is a linear process. This may seem obvious but we will see in the next lecture that this is certainly not true for the concept of variance.

So

$$E(Y) = E(3X+4) = 3E(X)+4$$
$$= 3\times 7+4 = \underline{\underline{25}}$$

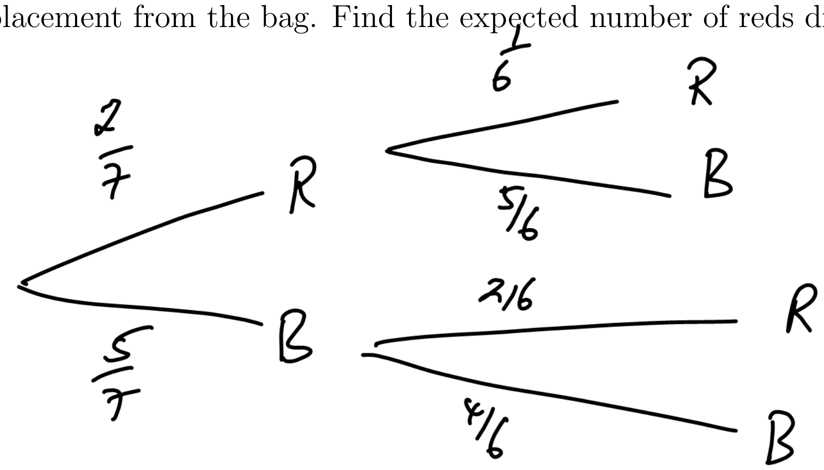
★ $E(Y) = 25$ ★

It is fortunate that we have the formula $E(Y) = E(3X + 4) = 3E(X) + 4$, for without it, we would need instead to re-work an expected value calculation for Y using $y_k = 3x_k+4$ in the transformed table:

y_k	10	13	16	19	22	25	28	31	34	37	40
$P(Y = y_k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Note we change the outcomes, NOT the probabilities.

Example 9: A bag contains 2 Red and 5 Black balls. Two balls are drawn without replacement from the bag. Find the expected number of reds drawn.



$P(RR) = \frac{2}{42}$

$P(RB) = \frac{10}{42}$

$P(BR) = \frac{10}{42}$

$P(BB) = \frac{20}{42}$

$X = \#$ Reds

x_k	0	1	2
$P(X=x_k)$	$\frac{20}{42}$	$\frac{20}{42}$	$\frac{2}{42}$

$$E(X) = 0 \times \frac{20}{42} + 1 \times \frac{20}{42} + 2 \times \frac{2}{42}$$

$$= \frac{24}{42} = \frac{4}{7}$$

$\star \quad \frac{4}{7} \quad \star$

Example 10: A random variable X has probability distribution:

x_k	a	b	c
$P(X = x_k)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Prove that if $E(X^2) = E(X)^2$ then $a = b = c$.

$$E(X) = a \times \frac{1}{3} + b \times \frac{1}{3} + c \times \frac{1}{3} = \frac{a+b+c}{3}$$

$$E(X)^2 = \frac{(a+b+c)^2}{9}$$

$E(X^2)$:

$Y = X^2$

x_k	a^2	b^2	c^2
$P(X=x_k)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$$\therefore E(X^2) = a^2/3 + b^2/3 + c^2/3 = \frac{a^2+b^2+c^2}{3}.$$

$$E(X^2) = E(X)^2$$

$$\Rightarrow \frac{a^2+b^2+c^2}{3} = \frac{(a+b+c)^2}{9}$$

$$\Rightarrow 3(a^2+b^2+c^2) = (a+b+c)^2 = (a^2+b^2+c^2) + 2ab + 2ac + 2bc$$

$$\Rightarrow 2(a^2+b^2+c^2) - 2ab - 2ac - 2bc = 0$$

$$\Rightarrow (a-b)^2 + (a-c)^2 + (b-c)^2 = 0 \quad \text{(nice trick!!)}$$

$$\Rightarrow a = b, \quad a = c, \quad b = c.$$

$\therefore$ $a = b = c$, as required. ★

It is clear from the above example that in general, $a \neq b \neq c$ and hence that, in general $E(X^2) \neq E(X)^2$.

Simple formulae like $E(aX + b) = aE(X) + b$ work nicely for linear transformations $Y = aX + b$. But finding $E(Y)$ for more complicated transformations like $Y = X^2$ will need to be done longhand. There are no shortcuts.

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 19

VARIANCE AND SPECIAL DISTRIBUTIONS

Milan Pahor

MATH1231 ALGEBRA
VARIANCE AND SPECIAL DISTRIBUTIONS

The **expected** value (also called the mean) of a discrete random variable X is given by

$$\mu = E(X) = x_1p_1 + x_2p_2 + \dots x_np_n = \sum_k x_kp_k$$

The **variance** $\mathrm{Var}(X)$ of a discrete random variable X which has mean $\mu = E(X)$ is given by

$$\sigma^2 = \mathrm{Var}(X) = (x_1-\mu)^2\,p_1 + (x_2-\mu)^2\,p_2 + (x_3-\mu)^2\,p_3 + \cdots + (x_n-\mu)^2\,p_n = \sum_k (x_k-\mu)^2\,p_k$$

OR $\qquad \mathrm{Var}(X) = E(X^2) - E(X)^2$

$$E(aX+b) = aE(X) + b$$

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$$

The Poisson distribution has probability distribution

$$P(X=k) = p_k = \frac{e^{-\lambda}\,\lambda^k}{k!}, \quad \text{for} \quad k = 0,1,2,3\dots$$

and a mean and variance of λ.

Variance Var(X)

To motivate the definition of variance let's take a look at two different data sets:

$$D_1 = \{11.8, 11.9, 12, 12.1, 12.2\}$$

and

$$D_2 = \{6, 9, 12, 15, 18\}$$

Then the average of the numbers in D_1 is $\dfrac{11.8+11.9+12+12.1+12.2}{5} = 12.$

The average of the numbers in D_2 is $\dfrac{6+9+12+15+18}{5} = 12.$

The two means are identical! But they are very different sets.

In D_1 all the numbers are tightly packed around the average 12. We say that D_1 has small **variance**.

But for D_2 the numbers are all over the place, either side of 12. We say that D_2 has large **variance** for the simple reason that the numbers *vary* more!

You will remember from the last lecture that the expected value $E(X)$ of a random variable X measures the centre of its distribution.

The variance $\mathrm{Var}(X)$ measures the **spread**. The higher the variance, the more chance of finding x_k's living a long way from the mean.

To find the variance of a random variable X you **must** first find the expected value $E(X)$, since variance is a measure of deviation from $E(X)$. The subsequent calculations are a bit gruesome.

Definition: The **variance** $\mathrm{Var}(X)$ of a discrete random variable X which has mean $\mu = E(X)$ is given by

$$\mathrm{Var}(X) = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + (x_3 - \mu)^2 p_3 + \cdots + (x_n - \mu)^2 p_n = \sum_k (x_k - \mu)^2 p_k.$$

Let's take a close look at this formula:

deviation from mean

(+)

$$\mathrm{Var}(X) = \sum_k (x_k - \mu)^2 p_k$$

mean

significance.

Observe that all the components in the calculation of variance are non-negative, so for any random variable X we have

$$\mathrm{Var}(X) \geq 0$$

The only way $\mathrm{Var}(X) = 0$ is if there is only one outcome with a probability of 1. That is, there is no variation in X.

Example 1: A discrete random variable X has the following distribution table.

x_k	1	2	3	4	5
p_k	0.05	0.2	0.5	0.2	0.05

a) Calculate the expected value $E(X)$.

b) Calculate the variance $\mathrm{Var}(X)$.

---

a) $E(X) = 1 \times .05 + 2 \times .2 + 3 \times .5 + 4 \times .2 + 5 \times .05$

$= 3$

b) To calculate the variance we add in another row to the table:

x_k	1	2	3	4	5
p_k	0.05	0.2	0.5	0.2	0.05
$(x_k - \mu)^2$	4	1	0	1	4

$(1-3)^2$ $(2-3)^2$ etc.

$\mathrm{Var}(X) = 4 \times .05 + 1 \times .2 + 0 \times .5 + 1 \times .2 + 4 \times .05$

$= 0.8$

★ $E(X) = 3$ and $\mathrm{Var}(X) = 0.8$ ★

Example 2: A discrete random variable Y has the following distribution table.

y_k	1	2	3	4	5
p_k	0.2	0.05	0.5	0.05	0.2

a) Calculate the expected value $E(Y)$.

b) Calculate the variance $\mathrm{Var}(Y)$.

---

a) $E(Y) = 1 \times .2 + 2 \times .05 + 3 \times .5 + 4 \times .05 + 5 \times .2$
$= 3$

b)

y_k	1	2	3	4	5
p_k	0.2	0.05	0.5	0.05	0.2
$(y_k - \mu)^2$	4	1	0	1	4

$\mathrm{Var}(Y) = 4 \times 0.2 + 1 \times .05 + 0 \times .5 + 1 \times .05 + 4 \times .2$

$= \underline{\underline{1.7}}$

★ $E(Y) = 3$ and $\mathrm{Var}(Y) = 1.7$ ★

Discussion: Why are the variance results for X and Y so different?

y_k's exist further from the mean with greater probabilities

★

Example 3: Two fair dice are tossed and the random variable X is the maximum of the two numbers which appear on the uppermost faces. Find the expected value and variance of X.

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

x_k	1	2	3	4	5	6
p_k	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

check: $\sum p_k = 1$ ✓

$$E(X) = 1\times\frac{1}{36} + 2\times\frac{3}{36} + 3\times\frac{5}{36} + 4\times\frac{7}{36} + 5\times\frac{9}{36} + 6\times\frac{11}{36}$$

$$= \frac{161}{36} \doteqdot 4.47.$$

$$\mathrm{Var}(X) = \left(1-\frac{161}{36}\right)^2\left(\frac{1}{36}\right) + \left(2-\frac{161}{36}\right)^2\left(\frac{3}{36}\right) + \left(3-\frac{161}{36}\right)^2\left(\frac{5}{36}\right) + \left(4-\frac{161}{36}\right)^2\left(\frac{7}{36}\right) + \left(5-\frac{161}{36}\right)^2\left(\frac{7}{36}\right) + \left(6-\frac{161}{36}\right)^2\left(\frac{11}{36}\right) = \frac{2555}{1296} \doteqdot 1.97.$$

★ $E(X) = \dfrac{161}{36} \approx 4.47$ and $\mathrm{Var}(X) = \dfrac{2555}{1296} \approx 1.97$ ★

Just as we denote the expected value of a random variable X by $E(X)$ or μ, so too the variance can be denoted by $\mathrm{Var}(X)$ or σ^2.

$$\sigma^2 = \mathrm{Var}(X) = (x_1-\mu)^2\,p_1+(x_2-\mu)^2\,p_2+(x_3-\mu)^2\,p_3+\cdots+(x_n-\mu)^2\,p_n = \sum_k (x_k-\mu)^2\,p_k$$

The fact that our units are being squared up in the calculation of the variance is sometimes of concern. To normalise the units we often also consider the standard deviation of a random variable, which is quite simply the square root of its variance.

$$\text{Standard deviation} = \sigma = \sqrt{\mathrm{Var}(X)}$$

Example 4: Find the standard deviation of the random variable X from the previous example.

$$\sigma = \sqrt{\frac{2555}{1296}}$$

★

The following fact gives us a cleaner and more efficient way of calculating variance and standard deviation.

Fact:

$$\mathrm{Var}(X) = E(X^2) - E(X)^2$$

Proof:

$$\sum_k (x_k-\mu)^2\,p_k =$$

$$= \sum_k (x_k^2 - 2x_k\mu + \mu^2)p_k.$$

$$= \sum_k x_k^2 p_k - 2\mu \sum_k x_k p_k + \mu^2 \sum_k p_k.$$

$$= E(X^2) - 2\mu(\mu) + \mu^2(1)$$

$$= E(X^2) - \mu^2 = E(X^2) - (E(X))^2$$ ★

Example 5: Recall in example 1 we showed that a random variable X with probability distribution

x_k	1	2	3	4	5
p_k	0.05	0.2	0.5	0.2	0.05

has variance given by $\sigma^2 = \frac{4}{5} = 0.8$

Redo this calculation using the new formula above, to find the variance and standard deviation of X.

We create a new table with the X^2 outcomes:

x_k	1	2	3	4	5
p_k	0.05	0.2	0.5	0.2	0.05
${x_k}^2$	1	4	9	16	25

Then $E(X) = 1 \times .05 + 2 \times .2 + 3 \times .5 + 4 \times .2 + 5 \times .05 = 3$

and

$E(X^2) = 1 \times .05 + 4 \times .2 + 9 \times .5 + 16 \times .2 + 25 \times .05 = 9.8$

Thus $\mathrm{Var}(X) = E(X^2) - E(X)^2 = 9.8 - (3)^2 = \underline{\underline{0.8}}$

and the standard deviation of X is

$\sigma = \sqrt{0.8}$.

★

Example 6: A random variable X has variance of 3 and $E(X^2) = 19$. Find the expected value of X.

$\mathrm{Var}(x) = E(x^2) - E(x)^2$

$3 = 19 - E(x)^2$

$\therefore (E(x))^2 = 16$

$E(x) = \underline{\underline{\pm 4}}$

★ $E(X) = \pm 4$ ★

We saw in the previous lecture that $E(aX + b) = aE(X) + b$. Variance behaves very, very differently under linear transformations!

Discussion: Why would we expect that $\mathrm{Var}(X + b) = \mathrm{Var}(X)$?

Shifting along does not impact on spread.

★

In general we have

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Example 7: A random variable X has mean $E(X)$ equal to 7 and variance $\mathrm{Var}(X)$ equal to 4. Let $Y = 3X + 12$. Find the expected value and variance of Y.

$$E(3x+12) = 3(E(x))+12 = 3\times 7+12 = \underline{33}$$

$$\mathrm{Var}(3x+12) = 9\,\mathrm{Var}(x)$$
$$= 9\times 4 = \underline{36}$$

★

Our final observation is that discrete random variables can most certainly have infinitely many possible outcomes, and that the sums defining $E(X)$ and $\mathrm{Var}(X)$ are then sums to infinity, just like in the Calculus strand!

Example 8: Suppose that a random variable X has a probability distribution given by

$$P(X = k) = p_k = \frac{e^{-7}\,7^k}{k!}, \qquad \text{for} \qquad k = 0, 1, 2, 3 \ldots$$

Note that we can't really show a table here as the table goes forever!

a) Prove that the total sum of all the probabilities is 1.

b) Calculate the probability that X is at least 2.

c) Prove that $E(X) = 7$.

It may be useful to recall from the calculus strand that

$$e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^k}{k!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

---

a) $\sum_k p_k = \sum_k e^{-7} 7^k/k! = e^{-7} \sum_k \frac{7^k}{k!} = e^{-7} e^7 = 1$ ✓

b) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(0 \text{ or } 1)$

$= 1 - \left\{ \frac{e^{-7} 7^0}{0!} + \frac{e^{-7} 7^1}{1!} \right\} = 1 - 8e^{-7}$

c) $E(X) = 0 \times p_0 + 1 \times p_1 + 2 \times p_2 + 3 \times p_3 + 4 \times p_4 + \cdots = 1 \times p_1 + 2 \times p_2 + 3 \times p_3 + 4 \times p_4 + \ldots$

$$= 1 \times \frac{e^{-7} 7^1}{1!} + 2 \times \frac{e^{-7} 7^2}{2!} + 3 \times \frac{e^{-7} 7^3}{3!} + 4 \times \frac{e^{-7} 7^4}{4!} + \ldots$$

$$= \frac{e^{-7} 7^1}{1!} + \frac{e^{-7} 7^2}{1!} + \frac{e^{-7} 7^3}{2!} + \frac{e^{-7} 7^4}{3!} + \ldots$$

$$= 7e^{-7} \underbrace{\left\{ 1 + \frac{7^1}{1!} + \frac{7^2}{2!} + \frac{7^3}{3!} + \ldots \right\}}_{e^7} = 7e^{-7} e^7 = 7.$$

★ *b)* $1 - 8e^{-7} = 0.9927$ ★

The above is a special case of the discrete Poisson distribution

$$P(X = k) = p_k = \frac{e^{-\lambda}\,\lambda^k}{k!}, \qquad \text{for} \qquad k = 0, 1, 2, 3 \ldots$$

which has a mean of λ and surprisingly, also a variance of λ.

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 20

BINOMIAL DISTRIBUTIONS AND THE SIGN TEST

Milan Pahor

MATH1231 ALGEBRA
BINOMIAL DISTRIBUTIONS AND THE SIGN TEST

Suppose we have independent repetitions of an experiment with probability p of success and probability $q = 1 - p$ of failure.

If X is the number of trials until the first success appears, then X follows the Geometric distribution:

$$P(X = k) = (1-p)^{k-1}p, \qquad k = 1, 2, 3, ...$$

with

$$E(X) = \frac{1}{p} \qquad \mathrm{Var}(X) = \frac{q}{p^2}.$$

If X is the number of successes in n trials then X follows the Binomial distribution $\mathrm{Bin}(n, p)$

$$P(X = k) = {}^nC_k \; p^k q^{n-k}, \qquad k = 0, 1, 2, 3, ..., n.$$

with

$$E(X) = np \;\text{ and }\; \mathrm{Var}(X) = npq.$$

We have created many of our own probability distribution tables in the previous two lectures. There are also however, many other famous and useful probability distributions in regular use, some with an infinite number of outcomes and others with a finite set of possible outcomes.

One such distribution was the Poisson distribution, presented at the end of the last lecture. Today we will look at two further important discrete distributions, the Geometric and the Binomial distributions.

Suppose that a biased coin when tossed has $P(H) = \dfrac{1}{3}$ and $P(T) = \dfrac{2}{3}$.

An experiment such as this with only two possible outcomes, is called a Bernoulli trial, and the repetition of independent Bernoulli trials is called a Bernoulli process. Two Bernoulli processes of interest in this lecture are Geometric and Binomial experiments.

Example 1: Suppose that a biased coin is tossed repeatedly with $P(H) = p = \dfrac{1}{3}$ and $P(T) = 1 - p = \dfrac{2}{3}$ for each of the tosses. We will treat head H as a success and tail T as a failure.

Let the random variable X be the number of tosses required for the first success (H).

a) Find $P(X = 4)$.

b) Find $P(X = k)$ for $k = 1, 2, 3, \ldots$.

c) Verify that the probability distribution in b) is valid.

a) $P(X=4) = P(TTTH)$

$= (2/3)^3(1/3) = \dfrac{8}{81}$

b) $P(X=k) = (2/3)^{k-1}(1/3)$

$k = 1, 2, 3, \ldots\ldots$

p_k.

c) Clearly $p_k \geq 0$.

$\sum p_k = \sum_{k=1}^{\infty} (2/3)^{k-1}(1/3)$

$= \frac{1}{3}\{1 + 2/3 + (2/3)^2 + \ldots\}$

$= \frac{1}{3}\left(\dfrac{1}{1-2/3}\right) = (1/3)\left(\dfrac{1}{(1/3)}\right)$

$= 1$ ✓✓

G.P.

$S_\infty = \dfrac{a}{1-r}$

$a = 1$

$r = 2/3$

$$\bigstar \qquad a) \quad \left(\frac{2}{3}\right)^3 \frac{1}{3} = \frac{8}{81} \quad b) \quad p_k = \left(\frac{2}{3}\right)^{k-1} \frac{1}{3}, \quad k = 1, 2, 3, \ldots \qquad \bigstar$$

The above example is a special case of the Geometric distribution:

Definition: Suppose that we have independent repetitions of an experiment with probability p of success and probability $q = 1 - p$ of failure, and we keep repeating the experiment until we get the first success.

If X is the number of trials required for the first success to appear, then X follows the Geometric distribution:

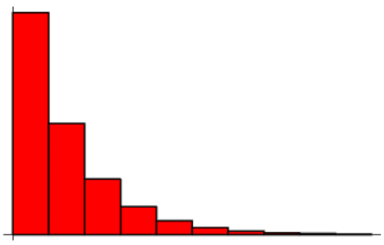
$$P(X = k) = (1 - p)^{k-1}p, \qquad k = 1, 2, 3, ...$$

and it is proven in your School notes that

$$E(X) = \frac{1}{p} \qquad \text{and} \qquad \text{Var}(X) = \frac{q}{p^2}.$$

$q = 1 - p$

The Geometric distribution, as you would expect, is heavily skewed to the left. In the above example it is very unlikely that it will take 8 tosses for the first head to appear.



$p(\text{failure}) = 3/4$
$p(\text{success}) = \frac{1}{4}$

Example 2: A new experimental missile has a probability of success only equal to $\frac{1}{4}$.

Let N denote the number of launches until the first missile success.

a) Find the mean and standard deviation of N.

b) Find $P(N = 6)$, the probability that the first success occurs on the sixth launch.

a) $E(x) = \frac{1}{p} = 4$, $V(x) = \frac{q}{p^2} = \frac{3/4}{\frac{1}{16}} = 12$ $\therefore \sigma = \sqrt{12}$

b) $\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)$

★ *a)* $E(X) = 4$ and $\sigma = \sqrt{12}$ *b)* $\left(\frac{3}{4}\right)^5 \left(\frac{1}{4}\right) \approx 0.059$ ★

Binomial Probability

Example 3: Suppose that a biased coin is tossed twelve times with $P(H) = p = \frac{1}{3}$ and $P(T) = q = 1 - p = \frac{2}{3}$ for each of the tosses. We will treat head H as a success and tail T as a failure.

In the twelve tosses, the most likely outcome is clearly 4 heads.

a) Could all the tosses be heads? Yes

b) Could all the tosses be tails? Yes

c) What is the probability of getting 5 heads in the 12 tosses?

d) Prove that the expected number of heads is 4. (This is intuitively clear, but lets see a proof).

---

c) ${}^{12}C_5 \left(\frac{1}{3}\right)^5 \left(2/3\right)^7$

d) It is clear from the above that the probability of k successes in 12 trials is given by

$$P(X = k) = {}^{12}C_k \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{12-k} \quad ; \; k = 0, 1, ..., 12.$$

$p = \frac{1}{3}, q = \frac{2}{3}$

Thus, to find the mean, we need to evaluate the sum

$$E(X) = \sum kp_k = \quad E(X) = \sum_{k=0}^{12} k \ ({}^{12}C_k) \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{12-k}.$$

Recall that the Binomial coefficient ${}^nC_k$ is often also written as $\begin{pmatrix} n \\ k \end{pmatrix}$.

Recall also the binomial theorem

$$(x + y)^{12} = \sum_{k=0}^{12} \left({}^{12}C_k\right) \ x^k y^{12-k}.$$

We can **partially** differentiate both sides with respect to x to obtain

$$12(x + y)^{11} = \sum_{k=0}^{12} \left({}^{12}C_k\right) \ kx^{k-1}y^{12-k}$$

and then multiply by x to obtain

$$12x(x + y)^{11} = \sum_{k=0}^{12} \left({}^{12}C_k\right) \ kx^k y^{12-k}$$

Now substituting $x = \left(\frac{1}{3}\right)$ and $y = \left(\frac{2}{3}\right)$ gives

$$12\left(\frac{1}{3}\right)(1)^{11} = \sum_{k=0}^{12} \left({}^{12}C_k\right) \ k \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{12-k}$$

Thus

$$4 = \sum_{k=0}^{12} k \ \left({}^{12}C_k\right) \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{12-k} = E(X).$$

★

So in general we have

Definition: The Binomial distribution is used for the situation when we have a total of n independent trials, in each of which there are only two possible outcomes with generic labels *Success* and *Failure*. We denote the chance of success on any trial by the constant p and the chance of failure by $q = 1 - p$.

If we let X be the number of successes in n trials the binomial distribution is given by

$$P(X = k) = {}^nC_k \; p^k q^{n-k},$$

for $0 \le k \le n$ and 0 otherwise.

If the random variable X follows a binomial distribution, we write $X \sim \text{Bin}(n, p)$, where the parameters n, p are as defined above.

It is shown in your School notes that

Fact: If $X \sim \text{Bin}(n, p)$, then $E(X) = np$ and $\text{Var}X = npq$.

Example 4: Legolas the archer can hit an Orc with with a probability of $\frac{1}{4}$.

Suppose that he shoots twenty times, and let the random variable X be the number of Orcs he hits in the twenty shots.

$p = \frac{1}{4}, \ n = 20$

a) Write down the expected value and variance of X.

b) Calculate $P(X = 7)$, that is, the probability of seven hits in twenty shots.

c) Calculate $P(X = 17)$, that is, the probability of seventeen hits in twenty shots. Compare c) ~~with b)~~.

d) What is the exact probability that Legolas hits at least 2 Orcs in twenty shots, that is, $P(X \ge 2)$.

a) $E(X) = np = 20\left(\frac{1}{4}\right) = 5$

$V(X) = npq = 20\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) = \frac{60}{16} = \frac{30}{8} = \frac{15}{4}$

b) $P(X = 7) = {}^{20}C_7 \left(\frac{1}{4}\right)^7 \left(\frac{3}{4}\right)^{13} = \underline{\underline{0.112}}$

c) $P(X = 17) = {}^{20}C_{17} \left(\frac{1}{4}\right)^{17} \left(\frac{3}{4}\right)^{3} \doteq \underline{\underline{\text{tiny}}}$

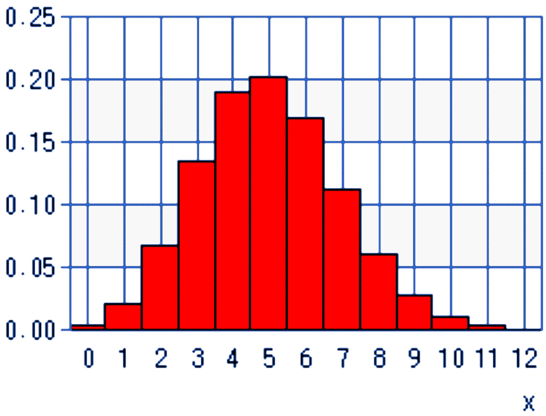
d) $P(X \geq 2) = 1 - P(X < 2)$

$= 1 - P(0 \text{ or } 1)$

$= 1 - \left\{ \left(\frac{3}{4}\right)^{20} + {}^{20}C_1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^{19} \right\}$

★ $a)\ E(X) = 5, \quad \mathrm{Var}(X) = \frac{15}{4} \quad b)\ {}^{20}C_7 \left(\frac{1}{4}\right)^7 \left(\frac{3}{4}\right)^{13} \approx 0.112 \quad c)\ {}^{20}C_{17} \left(\frac{1}{4}\right)^{17} \left(\frac{3}{4}\right)^3 \approx 0.000000028 \quad d)\ 1 - \left(\frac{3}{4}\right)^{20} - {}^{20}C_1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^{19}$ ★

Let's take a look at the above distribution.



Observe how the probability distribution is centered around the mean of 5 and roughly symmetric. There are two tails (on the distribution, not the Orcs), where the outcomes have little probability. Indeed outcomes above 12 are scarcely visible and have been omitted. The fact that events in the tails are extremely unlikely is exploited in the following section, Sign Tests.

Sign Tests

Example 5: Disappointed with his previous archery performance, Legolas studies the *Bows and Arrows* elf help manual. He soon announces to Gandalf that he has improved to the point that his probability of hitting an Orc is now equal to $\frac{1}{2}$.

Gandalf is doubtful, so runs a test, and Legolas hits 2 Orcs out of 10 in the next battle. Should Gandalf believe Legolas' claim of improvement?

This is a tricky situation, because **anything** is possible! However common practice is to initially accept the claim and then calculate the probability of the observed event or worse, to actually occur. We then to reject the claim if that probability falls below a level of significance, which in this course will be 5%.

In other words, if Legolas really has improved, he should hit more than 2 times out of 10. This performance certainly raises doubts, but how much doubt?

Let's assume that $P(\text{Hit})$ is now equal to 0.5, and that ten shots are taken.

Let the random variable X be the number of Orcs he hits in the ten shots.

We calculate $P(X \leq 2)$, which is the probability of the observed performance or worse. This is called a "tail probability" and gives us an idea how likely it is to find ourselves so deep in the tail or even further. This little tail probability always involves either the small left hand tail or right hand tail of the distribution, measured from the actual observation.

So $P(X \leq 2) = P(X = 0, 1, 2)$

$$= {}^{10}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10} + {}^{10}C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^9 + {}^{10}C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^8 \approx 0.054.$$

So, if Legolas **had** improved to level he claimed, the observed performance or worse would have occurred 5.4% of the time. Gandalf would have to grudgingly accept Legolas' claim at a level of significance of 5%, since his performance,though disappointing, has not quite intruded into the questionable bottom 5% tail.

The Single Sign Test

We can use the above ideas to test how reasonable a particular hypothesis is. We do this using a *sign test.*

The following example shows how the single sign test works.

Example: The following data are 15 measurements of moisture retention using a new sealing system. The previous system had a retention rate of 36 and the new system is expected to be higher. We wish to check if the new system is in fact superior.

$$37.5, \quad 35.2, \quad 37.3, \quad 36.0, \quad 36.8, \quad 39.8, \quad 37.4, \quad 35.3,$$
$$38.2, \quad 39.1, \quad 36.1, \quad 37.6, \quad 38.2, \quad 34.5, \quad 39.4.$$

To analyse the claim, we replace each score with the sign of the difference between the score and 36, including the possibility of this being 0. This gives the sequence

$$+ \quad - \quad + \quad 0 \quad + \quad + \quad + \quad - \quad + \quad + \quad + \quad + \quad + \quad - \quad + \,.$$

This gives 11 + signs.

We assume that there is no difference between the old system and the new system.
Then we would be equally likely to get + as − and the probability of + would be $\frac{1}{2}$.

So we have 14 differences of which 11 are +. This seems odd! It should be six out of 12? Let X be the number of plus signs. Then

$$\begin{aligned}
&P(X \geq 11) \\
&= {}^{14}C_{11}\left(\frac{1}{2}\right)^{11}\left(\frac{1}{2}\right)^{3} + {}^{14}C_{12}\left(\frac{1}{2}\right)^{12}\left(\frac{1}{2}\right)^{2} + {}^{14}C_{13}\left(\frac{1}{2}\right)^{13}\left(\frac{1}{2}\right)^{1} + {}^{14}C_{14}\left(\frac{1}{2}\right)^{14}\left(\frac{1}{2}\right)^{0} \\
&\approx 0.029.
\end{aligned}$$

This is less than 3% and very unlikely!

The assumption that there is no difference between the two systems is questionable at a level of significance of 5%.

This would provide strong evidence that the mean retention for the new system is greater than 36.

The Paired Sign Test.

The paired sign test can be used to compare two sets of data.

Example: The table below shows the hours of relief provided by two analgesic drugs in 12 patients suffering from arthritis. Is there any evidence that one drug provides longer relief than the other?

Case	Drug A	Drug B	B−A
1	2.0	3.5	+
2	3.6	5.7	+
3	2.6	2.9	+
4	2.6	2.4	−
5	7.3	9.9	+
6	3.4	3.3	−
7	14.9	16.7	+
8	6.6	6.0	−
9	2.3	3.8	+
10	2.0	4.0	+
11	6.8	9.1	+
12	8.5	20.9	+

If the drugs were equally effective, we would expect the difference in the hours of relief to be positive as often as it is negative. A result of nine out of twelve is borderline.

Let X be the number of times the difference $B - A$ is positive and assume that there is no difference between the drugs. Then $X \sim B(12, 0.5)$.

Now $P(X \geq 9) = \left\{\binom{12}{9} + \binom{12}{10} + \binom{12}{11} + \binom{12}{12}\right\}\left(\frac{1}{2}\right)^{12} \approx 0.07 \approx 7\%$

At a 5% level of significance we could **not** conclude that there is a difference between the drugs.

Note that the sign test is rather crude. It measures only the sign of the difference, and not the difference itself.

$${}^{12}C_9\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^3 + {}^{12}C_{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^2$$

$$+\,{}^{12}C_{11}\left(\frac{1}{2}\right)^{11}\left(\frac{1}{2}\right)^1 + {}^{12}C_{12}\left(\frac{1}{2}\right)^{12}\left(\frac{1}{2}\right)^0$$

$$= \left\{{}^{12}C_9 + {}^{12}C_{10} + {}^{12}C_{11} + {}^{12}C_{12}\right\}\left(\frac{1}{2}\right)^{12}$$

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 21

CONTINUOUS RANDOM VARIABLES

Milan Pahor

MATH1231 ALGEBRA
CONTINUOUS RANDOM VARIABLES

A function f is said to be a probability density function if

(I) $f \geq 0$ and

(II) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$

If a random variable X has p.d.f. f then

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

For a continuous random variable X with probability density function f

$$\mu = E(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

and

$$\sigma^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2\,f(x)\,dx = E(X^2) - E(X)^2.$$

We turn now to the issue of continuous random variables. Recall that

• Discrete random variables can only take on a countable number of distinct separate values. For example cars in a parking lot, hairs on a head, numbers on a die.

• Continuous random variables can take on any value within a range. For example height of a giraffe or length of a time interval.

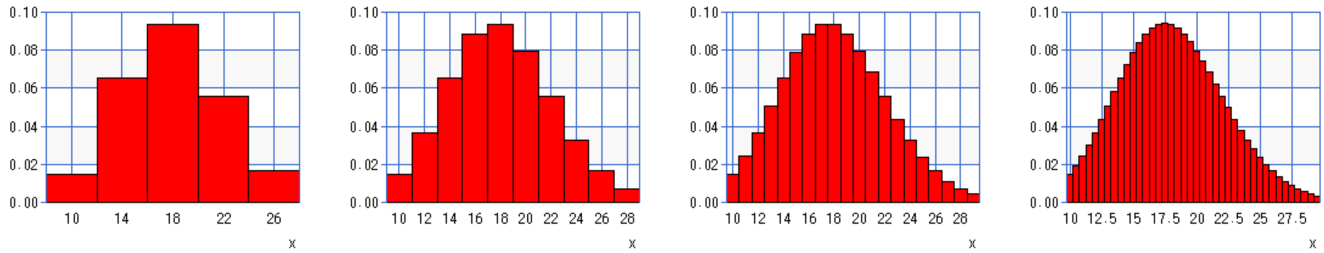
For discrete random variables we calculate a lot of sums. For continuous random variables it is integration! Yes, calculus will be involved.

For discrete random variables we consider tables of probabilities. On the other hand, for continuous random variables we use probability density functions.

A probability density function (also called a probability distribution function) is a non-negative function which gives us a feel for where the mass of the probabilities for X lie.

Probability density functions are unfortunately abbreviated to p.d.f's.

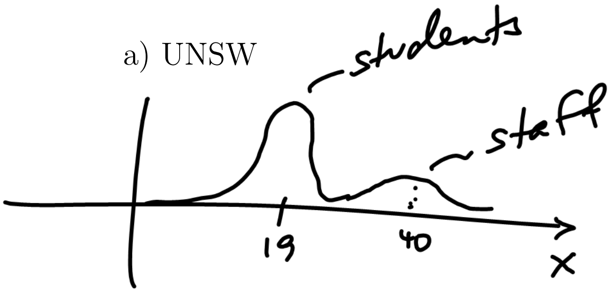
A p.d.f. is a little like a smooth version of a histogram:



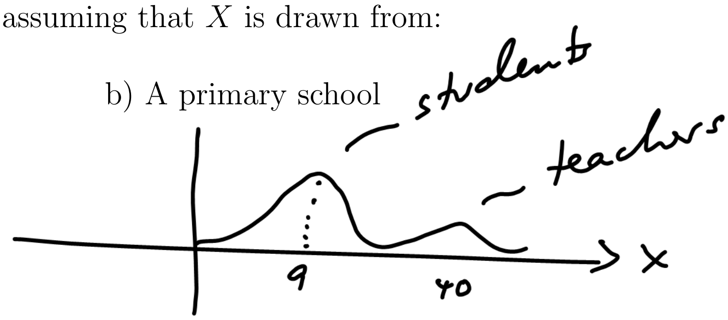
HISTOGRAM $\Longrightarrow$ $\Longrightarrow$ $\Longrightarrow$ $\Longrightarrow$ $\Longrightarrow$ $\Longrightarrow$ $\Longrightarrow$ **P.D.F**

Example 1: Suppose that the random variable X is the age of a group of people. Draw a possible probability density function of X assuming that X is drawn from:

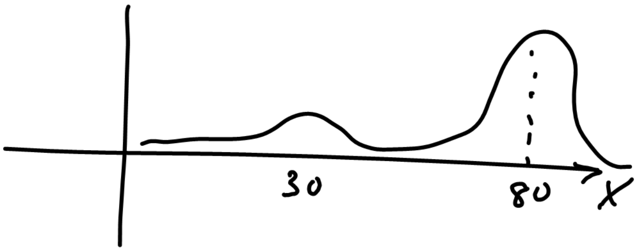
a) UNSW



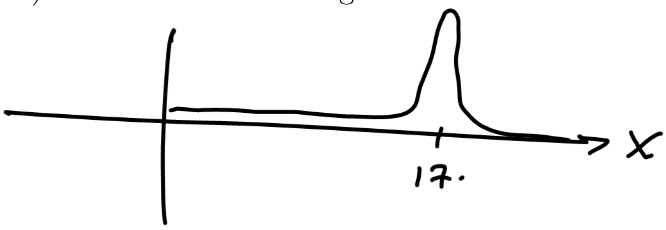
b) A primary school



c) An old age home.



d) A Year 12 class at high school.

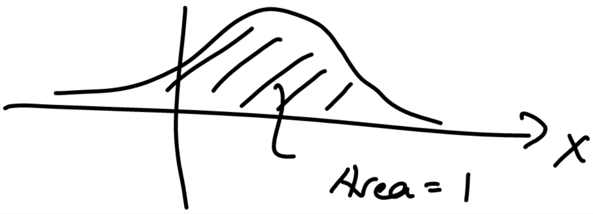


★

It is important to understand that a probability density function f does not need to be continuous or differentiable. It can be any shape or size and only needs to satisfy two critical conditions:

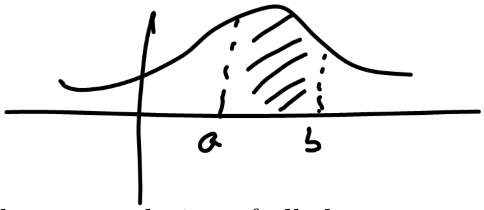
(I) $f \geq 0$ and

(II) $\int_{-\infty}^{\infty} f(x)\,dx = 1.$



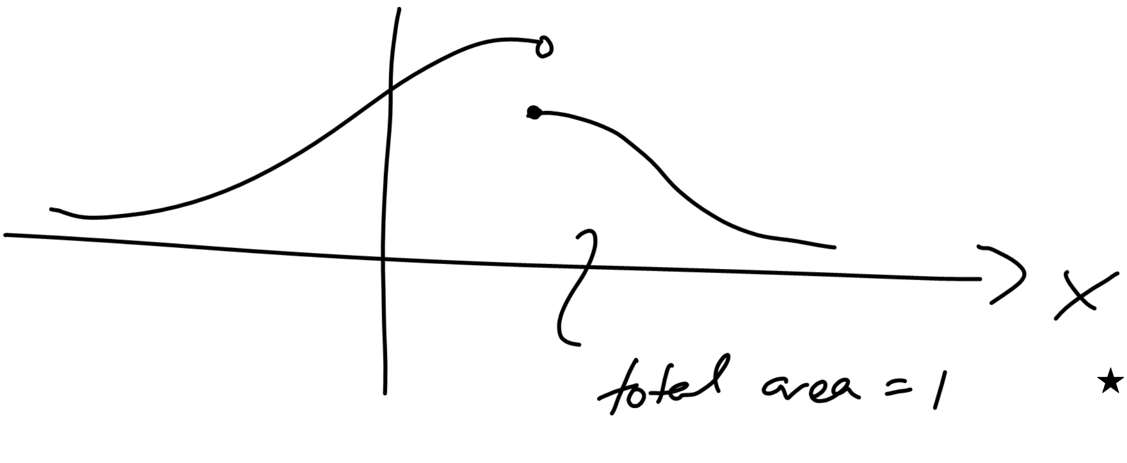
If a random variable X has p.d.f. f then

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$



That is, the single unit of area sitting under f is actually the accumulation of all the probability for X.... just like a histogram.

Discussion:



★

Example 2: A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \dfrac{x^3}{64} & 0 \leq x \leq 4 \\ 0 & \text{Otherwise.} \end{cases}$$

a) Sketch f.

b) Verify that f is a p.d.f.

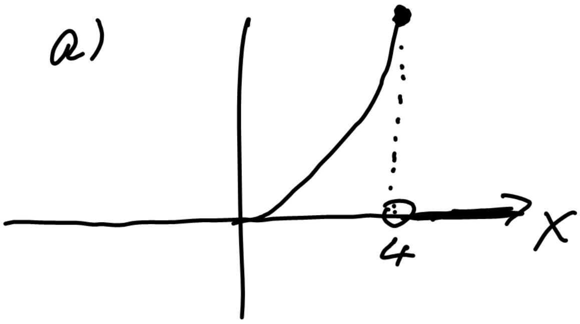
c) Find $P(0 \leq X \leq 1)$.

d) Find $P(X = 1)$.

e) Find $P(0 < X < 1)$.

f) Find $P(X \geq 3)$.

a)



b) i) $f \geqslant 0$

(ii) $\int_{-\infty}^{\infty} f(x)\,dx.$

$= \int_0^4 \frac{x^3}{64}\,dx = \left[\frac{x^4}{256}\right]_0^4$

$= \frac{4^4}{256} = 1.$

c) $P(0 \leq X \leq 1)$

$= \int_0^1 \frac{x^3}{64}\,dx = \left[\frac{x^4}{256}\right]_0^1$

$= \frac{1}{256}$

d) $P(X=1) = P(1 \leq X \leq 1)$

$= \int_1^1 f(x)\,dx = 0$!!!

e) $P(0 < X < 1)$

$= P(0 \leq X \leq 1)$

$= \frac{1}{256}$

f) $P(X \geqslant 3) = \int_3^{\infty} f(x)\,dx$

$= \int_3^4 \frac{x^3}{64} = \left[\frac{x^4}{256}\right]_3^4$

$= \frac{4^4}{256} - \frac{3^4}{256} = \frac{175}{256}$

★ $c)$ $\frac{1}{256}$ $d)$ 0 $e)$ $\frac{1}{256}$ $f)$ $\frac{175}{256}$ ★

Example 3: A continuous random variable X has a negative exponential density function given by

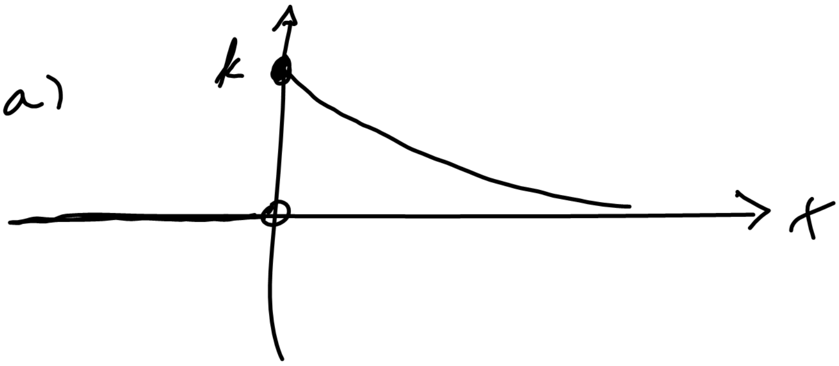
$$f(x) = \begin{cases} ke^{-2x} & x \geq 0 \\ 0 & \text{Otherwise.} \end{cases}$$

a) Sketch f.

b) Find k.

c) Find $P(3 \leq X \leq 4)$.

d) Find $P(X = 4)$.



b) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

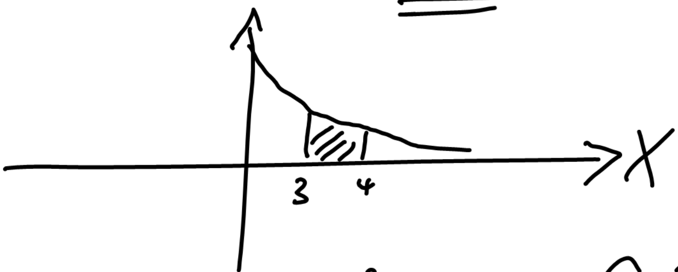
$\Rightarrow \int_0^{\infty} ke^{-2x}\,dx = 1 \Rightarrow k\int_0^{\infty} e^{-2x}\,dx = 1$

$\int_0^{\infty} e^{-2x}\,dx = \left[-\frac{1}{2}e^{-2x}\right]_0^{\infty} = \cancel{-\frac{1}{2}e^{-\infty}} - -\frac{1}{2}e^{-0}$

$= \frac{1}{2}.$

$\therefore\ k\left(\frac{1}{2}\right) = 1 \Rightarrow \underline{\underline{k = 2}}$

c)



$P(3 \leq X \leq 4) = \int_3^4 2e^{-2x}\,dx = \left[\frac{2e^{-2x}}{-2}\right]_3^4$

$= \left[-e^{-2x}\right]_3^4 = -e^{-8} - -e^{-6} = e^{-6} - e^{-8}$

d) 0

★ *b)* 2 *c)* $e^{-6} - e^{-8} \approx 0.00214$ *d)* 0 ★

For a continuous random variable X the concepts of expected value, variance and standard deviation are unchanged from the discrete case. We still have $\mu = E(X)$ as a measure of the center of the distribution, and $\sigma^2 = \mathrm{Var}(X)$ a measure of spread. The formulae are also similar except that we integrate rather than sum.

Definition: For a continuous random variable X with probability density function f

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

(measure of centre)

(centre of mass!)

and

$$\sigma^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx = E(X^2) - E(X)^2.$$

(measure of spread)

Students doing Physics will recognise the $E(X)$ formula as being that of the center of mass of a rod of varying density $f(x)$. The concepts are equivalent.

Example 4: A continuous random variable X has a probability density function given by

$$f(x) = \begin{cases} \dfrac{x^2}{63} & 3 \le x \le k \\ 0 & \text{o.w.} \end{cases}$$

~ otherwise

a) Sketch f.

b) Find k.

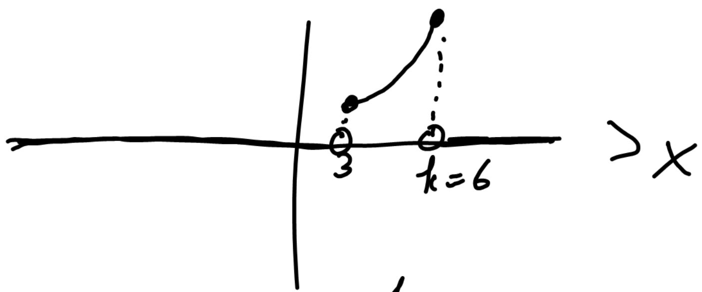
c) Find $E(X)$.

d) Find $\mathrm{Var}(X)$ using both formulae.

e) Find the standard deviation σ of X.

The calculation of some of the integrals in this example are left for homework. They are trivial, but horrible.

a)



b) $\displaystyle\int_3^k \frac{x^2}{63}\, dx = \left[\frac{x^3}{189}\right]_3^k = \frac{k^3}{189} - \frac{27}{189} = 1$

$k^3 - 27 = 189 \Rightarrow k^3 = 216 \Rightarrow k = 6$

c) $E(X) = \int_{-\infty}^{\infty} x f(x) dx.$

$$= \int_3^6 x \left(\frac{x^2}{63}\right) dx = \int_3^6 \frac{x^3}{63} dx.$$

$$= \left[\frac{x^4}{252}\right]_3^6 = \frac{6^4}{252} - \frac{3^4}{252}$$

$$= \frac{135}{28} \doteqdot 4.82$$

d) $Var(X) = \int_3^6 \left(x - \frac{135}{28}\right)^2 \left(\frac{x^2}{63}\right) dx = \ldots\ldots = \frac{2619}{3920}$

OR.

$$E(X^2) = \int_3^6 x^2 f(x) dx = \int_3^6 x^2 \left(\frac{x^2}{63}\right) dx$$

$$= \int_3^6 \frac{x^4}{63} dx = \ldots\ldots = \frac{7533}{315}.$$

$$\therefore Var(X) = E(X^2) - E(X)^2$$

$$= \frac{7533}{315} - \left(\frac{135}{28}\right)^2 = \frac{2619}{3920}$$

$$\sigma = \sqrt{\frac{2619}{3920}}$$

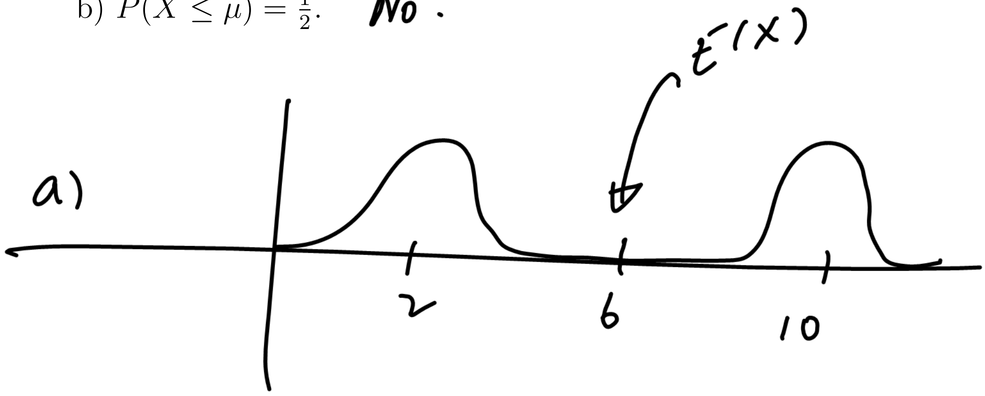
★ $b)$ $k = 6$ $c)$ $\mu = \frac{135}{28} \approx 4.82$ $d)$ $\sigma^2 = \frac{7533}{315} - \left(\frac{135}{28}\right)^2 = \frac{2619}{3920}$ $e)$ $\sigma = \sqrt{\frac{2619}{3920}}$ ★

Discussion: Suppose that X is a continuous random variable with p.d.f. f and expected value $\mu = E(X)$. Is it true that:

a) Most of the area under f is concentrated near μ. No.

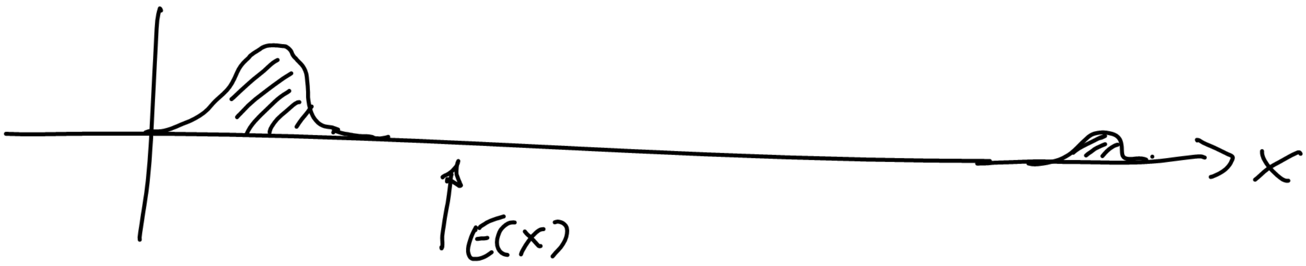
b) $P(X \leq \mu) = \frac{1}{2}$. No.

a)



$\therefore E(X) \cong 6$

b)



★

Our last comment is that the two formulae

$$E(aX + b) = aE(x) + b$$

and

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

still hold in the continuous case.

Note also that if $g(X)$ is a function of X then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

In particular

$$E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\, dx.$$

There are many different probability density functions but without doubt, the most useful continuous random variable is the Normal distribution. Our last two lectures will examine the Normal curve and its many applications.

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 22

THE STANDARD NORMAL DISTRIBUTION

Milan Pahor

MATH1231 ALGEBRA
THE STANDARD NORMAL DISTRIBUTION

If a continuous random variable X has probability density function f then the cumulative distribution function F of X is given by $$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt.$$
$$f = F'$$
The standard normal random variable is $Z \sim N(0,1)$. Probabilities $\Phi(z) = P(Z \le z)$ are extracted from the standard normal table.

In the last lecture we looked at probability density functions, which are essentially like continuous histograms. Recall that a function f is said to be a probability density function if

$$f \ge 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

If a continuous random variable X has probability density function f, then probabilities can be calculated via

$$P(a \le X \le b) = \int_{a}^{b} f(x)\,dx.$$

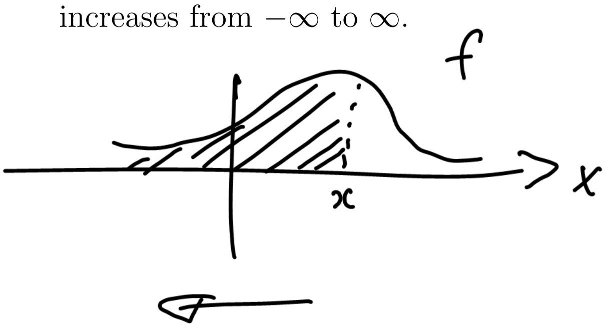
We now introduce a new type of function F called a cumulative distribution function.

Definition: If a random variable X has probability density function f then the cumulative distribution function F of X is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt.$$

Note in the above definition that, since we have an x as a limit, we need to use the dummy variable t in the integral.

The cumulative distribution function simply sweeps up all available probability as x increases from $-\infty$ to ∞.



Example 1: Suppose that a continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} k & 4 \leq x \leq 10 \\ 0 & \text{Otherwise.} \end{cases}$$

a) Find the value of k.

b) Sketch the p.d.f. f.

c) Find and sketch the cumulative distribution function F of X.

d) Calculate $P(X \leq 8)$.

e) Calculate $P(X \geq 5)$.

---

a) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

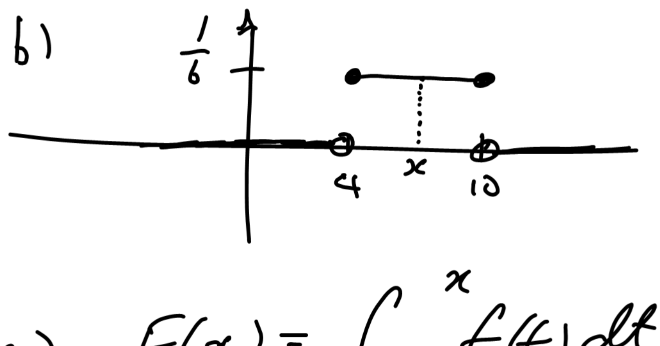
$\Rightarrow \int_{4}^{10} k\,dx = 1$

$\Rightarrow [kx]_{4}^{10} = 1$

$\Rightarrow 10k - 4k = 1 \Rightarrow 6k = 1$

$\Rightarrow k = \frac{1}{6}$

b)



c) $F(x) = \int_{-\infty}^{x} f(t)\,dt$

If $x < 4 \rightarrow F(x) = 0$

If $x > 10 \rightarrow F(x) = 1$

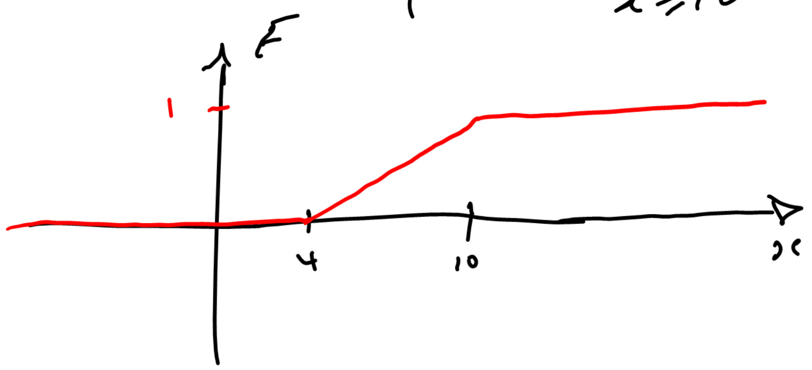
If $4 \leq x \leq 10$

$F(x) = \int_{-\infty}^{x} f(t)\,dt$

$= \int_{4}^{x} \frac{1}{6}\,dt = \left[\frac{1}{6} t\right]_{4}^{x}$

$= \frac{1}{6}x - \frac{1}{6}4 = \frac{x}{6} - 2/3$

$$F(x) = \begin{cases} 0 & x < 4 \\ x/6 - 2/3 & 4 \leq x \leq 10 \\ 1 & x \geq 10 \end{cases}$$



d) $P(X \leq 8) = F(8)$

$= 8/6 - 2/3 = \frac{4}{3} - 2/3 = 2/3$

e) $P(X \geq 5) = 1 - P(X < 5)$

$= 1 - F(5)$

$= 1 - \left(\frac{5}{6} - \frac{2}{3}\right)$

$= 1 - \frac{1}{6} = \frac{5}{6}.$

★ a) $k = \dfrac{1}{6}$ c) $F(x) = \begin{cases} 0 & x < 4 \\ \dfrac{x}{6} - \dfrac{2}{3} & 4 \leq x \leq 10 \\ 1 & x > 10 \end{cases}$ ★

You will observe from the above sketches that the cumulative distribution function F is continuous, even though the probability density function f isn't continuous. This is because the probability is always picked up infinitesimally. You can't get a sudden lump of probability!

Discussion: Why will every cumulative distribution function F be an increasing function, rising from zero on the extreme left, to 1 on the extreme right?

F is increasing because it is accumulating area

extreme left $\Rightarrow$ no area to the left

extreme right $\Rightarrow$ all the area is to the left!

★

As you would well imagine, if a continuous random variable X has a p.d.f. f and a cumulative distribution function F then

$$f = F'.$$

This is just the first fundamental theorem of calculus! That is, F comes from f via integration, so f comes from F through differentiation.

Example 2: A continuous random variable X has a cumulative distribution function F given by

$$F(x) = \begin{cases} 0 & x < 3 \\ \\ \dfrac{x^2}{40} - \dfrac{9}{40} & 3 \le x \le 7 \\ \\ 1 & x > 7. \end{cases}$$

a) Sketch F.

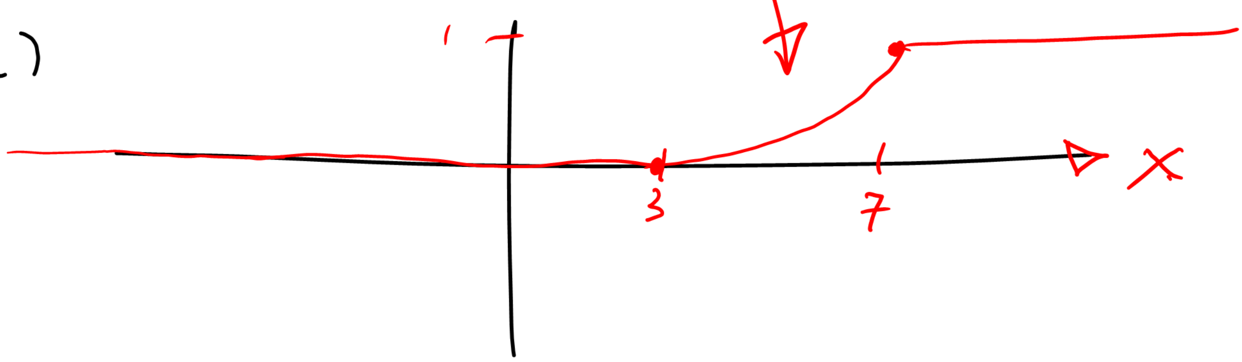
b) Find $P(X \geq 5)$.

c) $P(4 < X \leq 6)$.

d) Find and sketch the probability density function f.

e) Check that p.d.f. f found in d) is valid.

a)



b) $P(X \geq 5) = 1 - P(X < 5)$

$= 1 - F(5)$

$= 1 - \left\{ \frac{25}{40} - \frac{9}{40} \right\}$

$= 1 - \frac{16}{40} = \frac{24}{40} = \frac{12}{20} = 3/5$

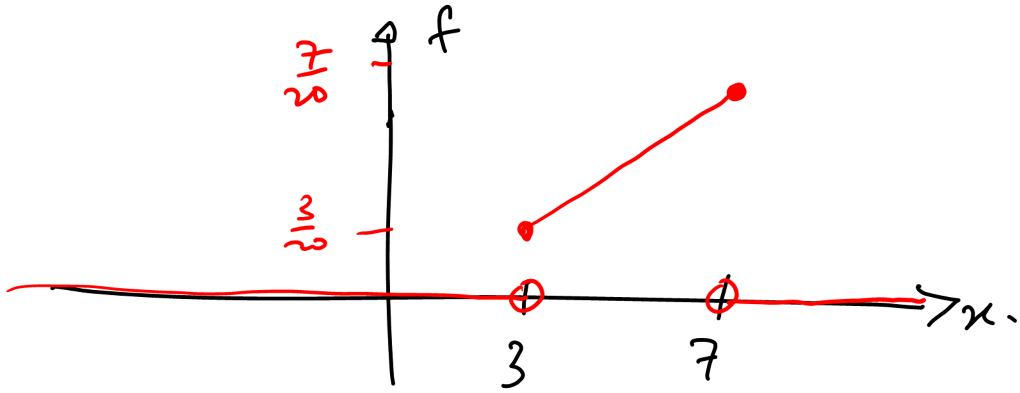
c) $P(4 < X \leq 6) = P(X \leq 6) - P(X \leq 4)$

$= F(6) - F(4)$

$= \left(\frac{36}{40} - \cancel{\frac{9}{40}} \right) - \left(\frac{16}{40} - \cancel{\frac{9}{40}} \right) = \frac{1}{2}$

d) $f = F'$

$$f(x) = \begin{cases} 0 & x < 3 \\ \dfrac{x}{20} & 3 \leq x \leq 7 \\ 0 & x > 7 \end{cases}$$



e) $f \geqslant 0$ ✓

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_3^7 \frac{x}{20}\,dx = \left[\frac{x^2}{40}\right]_3^7$$

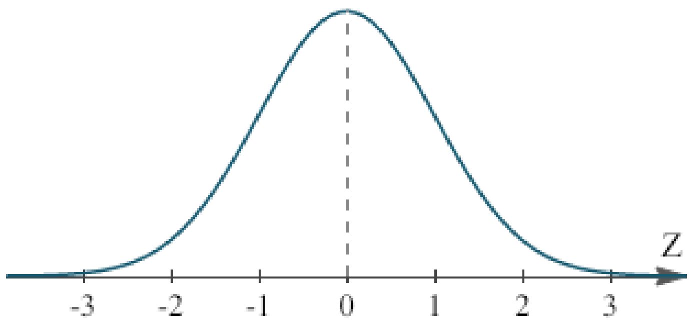
$$= \frac{49}{40} - \frac{9}{40} = \frac{40}{40} = 1 \quad \checkmark\checkmark$$

$\therefore$ Total area $= 1$ as required.

★

The Standard Normal Distribution

Without doubt, the most important probability density function we encounter in the real world is the standard normal distribution $Z \sim N(0, 1)$. The normal distribution has the following beautifully balanced and elegant look:



$N(0, 1) \sim Z$

μ σ^2

When we say $Z \sim N(0, 1)$ we are saying that Z is normally distributed with a mean of 0 and a variance of 1.

You will see in second year that almost all statistical issues and questions eventually find their way back to the normal distribution. Sums and averages of almost all populations look more and more normal as the size of the population increases.

Tragically the actual equation of the normal curve is horrendous

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$$

which completely stops us from calculating probabilities using integration. Instead we have to use a standard normal table of probabilities, attached at the end of the lecture.

This table supplies us with

Phi of z

$$\Phi(z) = P(Z \leq z)$$

Thus it actually gives us the cumulative probability distribution function of Z.

Note that

• z is only measured to 2 decimal places and the probabilities to 4 decimal places.

• The extreme left hand column in the standard normal table displays the $z-$score to one decimal place, and the top row then takes you to a second decimal place, if needed.

• The table **only** looks left.....like any cumulative distribution function.

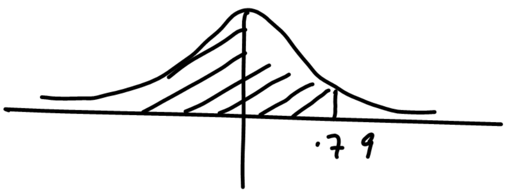
• There are slightly different presentations of this table in different sources.

• All most probabilities are exhausted by the time you hit 3 standard deviations from the mean.

Let's take a look at how the table is used. It's easy:

Example 3: Suppose that $Z \sim N(0,1)$. Use the standard normal table to calculate each of the following probabilities. In each case, sketch the region of interest:

a) $P(Z \leq 0.79)$. $= \Phi(0.79) = \underline{\underline{0.7852}}$



b) $P(Z = 0.79)$.

$= 0$

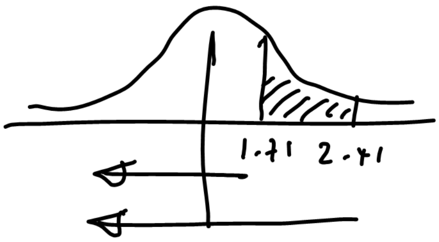
c) $P(Z < 0.79)$. $= \Phi(.79) = 0.7852$

d) $P(Z \geq 2.47)$. $= 1 - P(Z < 2.47) = 1 - \Phi(2.47)$

$= 1 - 0.9932 = \underline{\underline{0.0068}}$

e) $P(Z \leq -1.24)$. $= \Phi(-1.24) = \underline{0.1075}$

f) $P(1.71 \leq Z \leq 2.41)$ $= \Phi(2.41) - \Phi(1.71)$



$= 0.9920 - 0.9564 = 0.0356$

★ *a*) 0.7852 *b*) 0 *c*) 0.7852 *d*) 0.0068 *e*) 0.1075 *f*) 0.0356 ★

We also need to able to use the table backwards, looking up probabilities in the main body of the table to produce z scores, $-3 < z < 3$.

Example 4: Suppose that $Z \sim N(0,1)$.

Sketch an appropriate region under the p.d.f of Z and use the standard normal table to find α such that:

a) $P(Z \leq \alpha) = 0.9$.

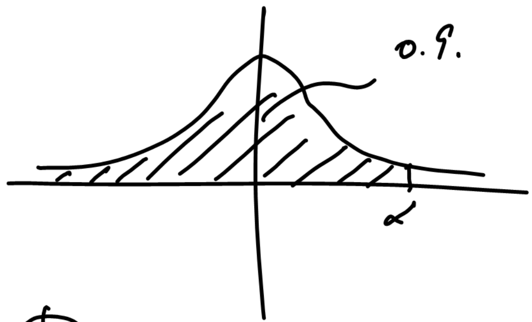
b) $P(Z \geq \alpha) = 0.01$.

c) $P(Z \leq \alpha) = 0.0735$.

---

a) Note that 0.9 does not appear exactly in the main body of the table. We simply pick off the nearest table entry.

a) $\alpha = 1.28$

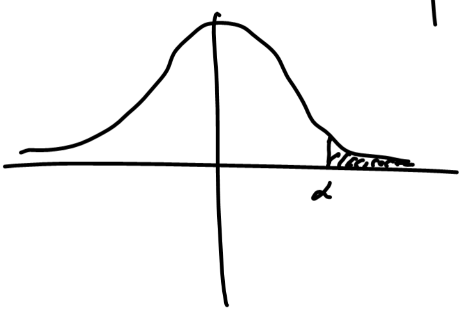


b) $P(Z \geq \alpha) = 0.01$

$\therefore 1 - P(Z < \alpha) = 0.01$

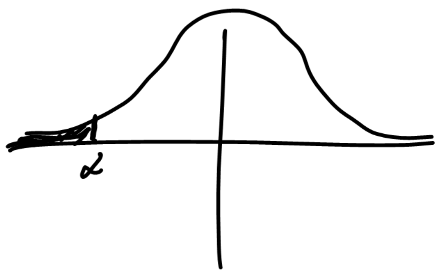
$0.99 = P(Z < \alpha)$

$\therefore \alpha = 2.33.$



c) $P(Z < \alpha) = 0.0735$

$\therefore \alpha = -1.45$



★ $a)$ 1.28 $b)$ 2.33 $c)$ -1.45 ★

We clearly have a problem. It's great to have a normal table with a mean of 0 and a variance of 1. But what if our normal distribution has a different mean or a different variance. How do we calculate probabilities then?

This is the content of our last lecture in this course.

Standard normal probabilities $P(Z \leq z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

School of Mathematics and Statistics

Math1231 Mathematics 1B

ALGEBRA LECTURE 23

GENERAL NORMAL DISTRIBUTIONS

Milan Pahor

MATH1231 ALGEBRA
GENERAL NORMAL DISTRIBUTIONS

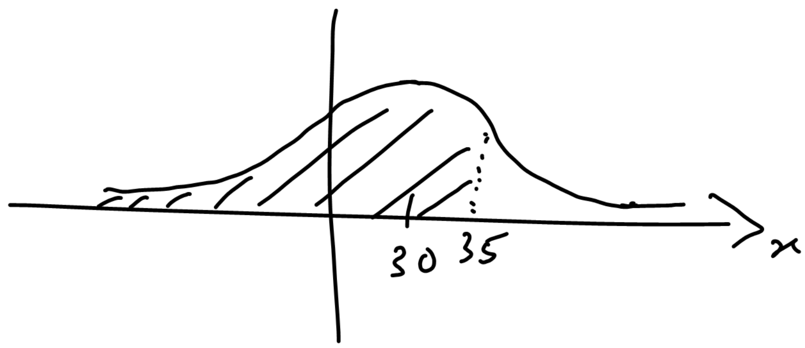
$$X \sim N(\mu, \sigma^2) \implies \frac{X - \mu}{\sigma} \sim N(0, 1)$$

If a **discrete** random variable X follows a binomial distribution with n trials and a probability of success of p, then X can be approximated by the **continuous** random variable $Y \sim N(np, npq)$.

Suppose that a random variable X is normally distributed with a mean of 30 and a variance of 9. We write

$$X \sim N(30, 9).$$

So X is centered about $x = 30$ and a variance larger than 1 means that the distribution is a bit more spread out than for $Z \sim N(0, 1)$. So the density function for X would like like:



★

Now consider the question: Find $P(X \leq 35)$. Sounds easy!

It is clear from the sketch that the answer should be close to 1. All we need to do is look up 35.00 in the table.........but the table stops at 2.99?

Our trick is to use a transformation to convert the problem back over to $Z \sim N(0, 1)$, so that the standard normal table can be used. Otherwise we would need a million different tables!

Fact:

$$X \sim N(\mu, \sigma^2) \implies \frac{X-\mu}{\sigma} \sim N(0,1)$$

Proof: Let's check that this transformation does indeed produce a mean of zero and a variance of 1. Assume that X has a mean of μ and a variance of σ^2. Recall that

$$E(aX+b) = aE(X)+b \quad \text{and} \quad \mathrm{Var}(aX+b) = a^2\mathrm{Var}(X).$$

So

$$E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}\left(E(X)-E(\mu)\right) = \frac{1}{\sigma}(\mu-\mu) = 0.$$

$$\mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \mathrm{Var}\left(\frac{X}{\sigma}-\frac{\mu}{\sigma}\right) = \mathrm{Var}\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1.$$

It is shown in your printed notes that the transformed random variable is still normal in shape.

★

So $N(\mu, \sigma^2)$ can be transformed to $N(0,1)$ by subtracting the mean and dividing by the standard deviation. Therefore for $X \sim N(30, 9)$:

$$P(X \le 35) = P\left(\frac{X-30}{3} \le \frac{35-30}{3}\right)$$

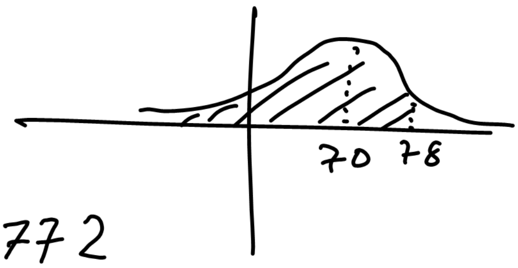
$$= P\left(Z \le \frac{5}{3}\right) = \Phi(1.67) = 0.9525$$

(table)

★ $\Phi(1.67) = 0.9525$ ★

Example 1: Suppose that $X \sim N(70, 16)$. Evaluate each of the following probabilities, first sketching an appropriate region, and then using the standard normal table.

a) $P(X \le 78)$.



$$= P\left(\frac{X-70}{4} \le \frac{78-70}{4}\right)$$

$$= P(Z \le 2) = \Phi(2) = \underline{0.9772}$$

b) $P(X = 78)$.

$$= 0$$

c) $P(X < 68)$.

$$= P\left(\frac{X-70}{4} < \frac{68-70}{4}\right) = P\left(Z < -\tfrac{1}{2}\right)$$

$$= \Phi(-0.5)$$

d) $P(X \geq 60)$.

$$= 1 - P(X<60) = 1 - P\left(\frac{X-70}{4} < \frac{60-70}{4}\right)$$

$$= 1 - P(Z < -2.5) = 1 - \Phi(-2.5)$$

e) $P(66 \leq X \leq 75)$ $= P\left(\frac{66-70}{4} < \frac{X-70}{4} < \frac{75-70}{4}\right)$

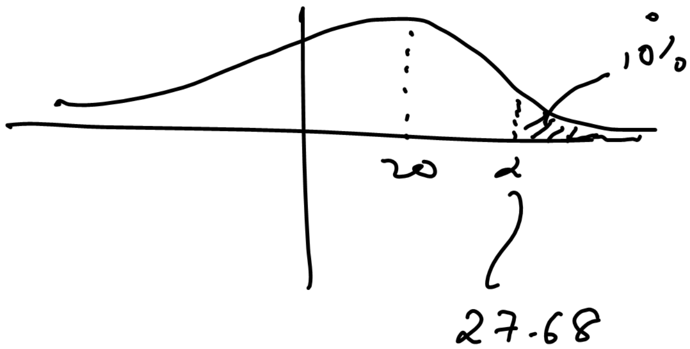
$$= P(-1 < Z < 1.25) = P(Z < 1.25) - P(Z < -1)$$

$$= \Phi(1.25) - \Phi(-1).$$

★

We still need to able to use the table backwards!

Example 2: Suppose that $X \sim N(20, 36)$. Find the value α of X exceeded by only 10% of the population. (This is called the upper 10% percentile.)



$$P(X < \alpha) = 0.9$$

But

$$P(Z < 1.28) = 0.9$$

$$\therefore 1.28 = \frac{X-20}{6}$$

$$\therefore X - 20 = 6 \times 1.28$$

$$\Rightarrow X = \underline{20 + 6 \times 1.28} \fallingdotseq 27.68$$

★ $\alpha = 20 + 6 \times 1.28 = 27.68$ ★

μ σ^2

Example 3: Suppose that the time (in minutes) taken to be served at a particular bank is normally distributed with a mean of 15 minutes and a variance of 4 minutes.

a) Find the probability that a random individual is served in less than 12 minutes.

b) Find the probability of being served in between 14 minutes and 17 minutes.

c) Find the quickest 1% of serving times.

a) $X \sim N(15, 4)$

$$P(X < 12) = P\left(\frac{X-15}{2} < \frac{12-15}{2}\right)$$

$$= P(Z < -1.5) = \Phi(-1.5)$$

$$= 0.0668$$

b) $P(14 \leq X \leq 17)$

$$= P\left(\frac{14-15}{2} < \frac{X-15}{2} < \frac{17-15}{2}\right) = P\left(-\frac{1}{2} \leq Z \leq 1\right)$$

$$= \Phi(1) - \Phi\left(-\frac{1}{2}\right) = 0.8413 - 0.3085$$

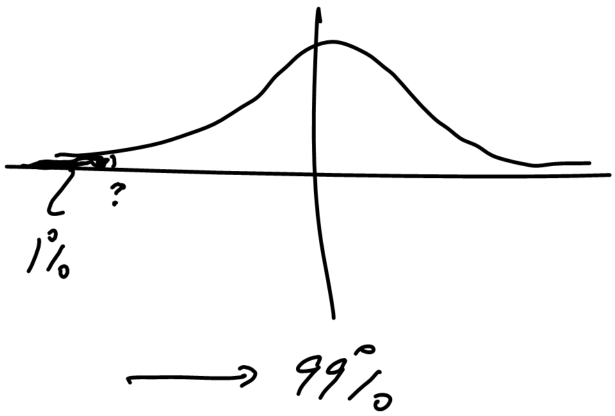
$$= \underline{0.5328}$$ ✓

c) $P(Z < -2.33) = 0.01$

$$\therefore \frac{X-15}{2} = -2.33$$

$$\Rightarrow X = 15 - (2.33)(2)$$

$$= 10.34 \text{ minutes}$$



★ *a)* .0668 *b)* .5328 *c)* $15 - 2 \times 2.33 = 10.34$ minutes ★

Normal Approximation to the Binomial Distribution

Example 4: Suppose that a biased coin with $P(H) = 0.8$ is tossed 50 times. What is the binomial probability of obtaining less than or equal to 42 heads? On average we'd expect 40 heads in 50 tosses, so this outcome seems quite likely.

But this is too much work, no matter what we do!

The exact answer is $P(X \leq 42) =$

$${}^{50}C_{42}\,(0.8)^{42}\,(0.2)^{8} + {}^{50}C_{41}\,(0.8)^{41}\,(0.2)^{9} + {}^{50}C_{40}\,(0.8)^{40}\,(0.2)^{10} + {}^{50}C_{39}\,(0.8)^{39}\,(0.2)^{11} + \cdots + {}^{50}C_{0}\,(0.8)^{0}\,(0.2)^{50}$$

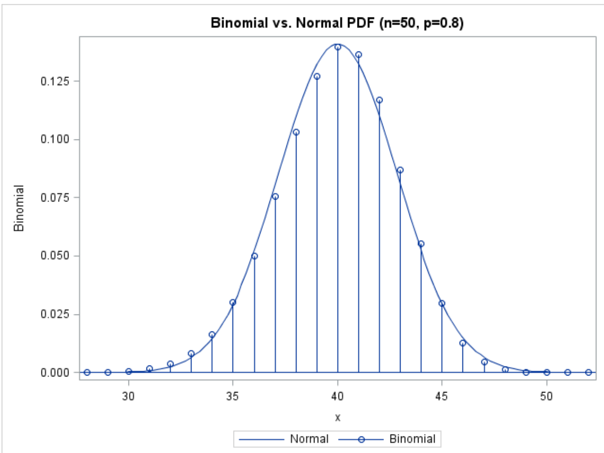
This is over forty calculations and even if we go to the complement the workload is still heavy! In situations like this where we have too many Binomial calculations to perform, there is a trick:

Let the Binomial discrete variable X count the number of heads in 50 tosses.

Recall that $E(X) = np = 50 \times 0.8 = 40$ and that $\text{Var}(X) = npq = 50 \times 0.8 \times 0.2 = 8$.

Now consider instead the **continuous** random variable $Y \sim N(40, 8)$ with the same mean and variance as X.

Placing the p.d.f. of Y over the histogram of X we see that they match almost perfectly! But there is one little correction needed:



Example 5: Calculate $P(X = 42)$ and P(Y=42).

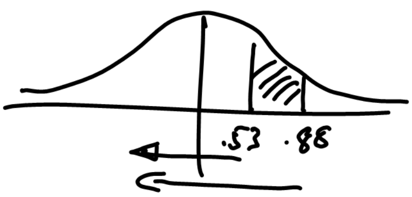
$${}_{50}C_{42}(.8)^{42}(.2)^{8} = P(X=42)$$

$$= \underline{0.1169}$$

$$P(Y=42) = P(41.5 \leq Y \leq 42.5)$$

$$= P\left(\frac{41.5-40}{\sqrt{8}} \leq \frac{Y-40}{\sqrt{8}} \leq \frac{42.5-40}{\sqrt{8}}\right)$$

★ $P(X = 42) = {}^{50}C_{42}\,(0.8)^{42}\,(0.2)^{8} = 0.1169,\quad P(Y = 42) = \Phi(0.88) - \Phi(0.53) = 0.1087 \approx 0.1169$ ★



$$= P(0.53 \leq Z \leq 0.88)$$

$$= \Phi(0.88) - \Phi(0.53) = 0.1087 \cong 0.1169$$

continuous Y

discrete X

So let's use the continuous Normal approximation to the discrete Binomial problem:

$P(X \leq 42) \approx P(Y \leq 42.5) = P\left(\frac{Y-40}{\sqrt{8}} \leq \frac{42.5-40}{\sqrt{8}} \right)$

$= P(Z \leq 0.88) = \Phi(0.88) = \underline{0.8106}$ ★ $\Phi(0.88) = 0.8106$ ★

Discussion: Why do we correct to 42.5 rather than use 42?

$41.5 \leftarrow 42 \rightarrow 42.5$

stretches $\frac{1}{2}$ unit left and right.

★

Note that we **only** correct the Normal calculations when we are approximating a discrete random variable. We never make corrections for straight up simple normal questions. The correction will always **add in** a little bit of extra (missed) probability.

> **Fact:** If a **discrete** random variable X follows a binomial distribution with n trials and a probability of success of p, then X can be approximated by the **continuous** random variable $Y \sim N(np, npq)$. $q = 1-p$

Example 6: Suppose that 2000 digits are randomly selected from the set of ten digits $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Find the probability that the digit 7 appears less than or equal to 190 times.

We would expect around 200 occurrences of a 7, so 190 seems reasonable. But we would need to sum up 190 Binomial probabilities! We let X count the number of occurrences of a 7 in the 2000 digits and move to the normal approximation instead.

We have $p = \frac{1}{10}$ and $n = 2000$. So

$E(X) = np = 2000 \times \frac{1}{10} = 200$ and $\text{Var}(X) = npq = 2000 \times \frac{1}{10} \times \frac{9}{10} = 180.$

So X can be approximated by the continuous normal random variable $Y \sim N(200, 180)$.

Then $P(X \leq 190) =$

$$P(Y \leq 190.5) \approx P\left(\frac{Y-200}{\sqrt{180}} \leq \frac{190.5-200}{\sqrt{180}}\right) = P(Z \leq -0.71) = \Phi(-0.71) = 0.2389.$$

★

Example 7: In a certain Scandinavian country every purchased car is either a Volvo or a Saab. Historically Volvo's dominate sales with 70% of the market. The Saab company runs an advertising campaign in order to increase its market share from 30%.

At the conclusion of the advertising campaign a random survey of 300 sales was conducted and it was found that 111 were Saabs instead of the expected 90.

At a level of significance of 5 percent, determine whether or not the survey results indicate that the advertising campaign was successful.

We begin by assuming that the campaign had no impact. That is, it is still true that only 30% of car sales are Saabs.

Let the discrete Binomial random variable X count the number of Saabs sold in 300 random purchases.

Then $E(X) = np = 300 \times 0.3 = 90$ and $\mathrm{Var}(X) = npq = 300 \times 0.3 \times 0.7 = 63$.

The probability of witnessing such a probability tail in the survey results is then $P(X \geq 111)$. We approximate the discrete X with continuous $Y \sim N(90, 63)$.

$$P(X \geq 111) \approx P(Y \geq 110.5) = P\left(\frac{Y - 90}{\sqrt{63}} \geq \frac{110.5 - 90}{\sqrt{63}}\right) = P(Z \geq 2.58)$$

$$= 1 - P(Z \leq 2.58) = 1 - \Phi(2.58) = 1 - 0.9951 = 0.0049.$$

This is only $\frac{1}{2}$ %, which is very unlikely!

We could safely conclude from the survey that Saab sales had in fact improved.

SOME FINAL INFORMATION

1. Please check online that all your marks are recorded correctly.

2. Read the school pages on additional assessment/special consideration so that you are fully aware of the rules that apply.

3. Past papers are on Moodle.

4. Make sure that you are aware of the format and date of the final exam. If in any doubt please consult the Moodle page or contact the first year office.

5. Please take the time to complete all online surveys regarding the administration and teaching of the course.

6. Check Moodle for consultation options during stuvac.

Good Luck!

Milan Pahor

Standard normal probabilities $P(Z \leq z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986