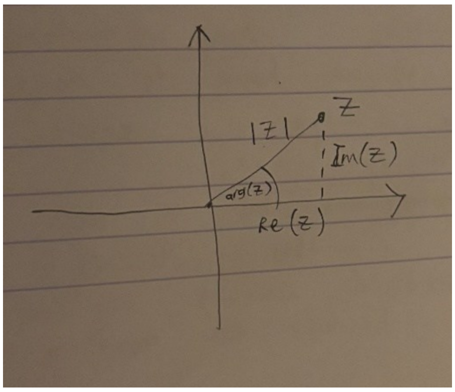
Lecture 1:

The course starts with a review of complex number properties from A level further maths which is because we will later look at vectors and matrices with complex number entries. There are two new things that were proven in lecture 1.

For any complex number z=a+bi, $\frac{1}{z} = \frac{a-bi}{a^2+b^2} = \frac{a-bi}{|a|^2}$, and this can be easily verified. Also, $z\bar{z} = |z|^2$, proven because the modulus and arguments agree.

For any complex number z, $\sin(\arg(z)) = \frac{Im(z)}{|z|}$ and $\cos(\arg(z)) = \frac{Re(z)}{|z|}$. This is immediate from drawing a little diagram (image below), as it falls right from the o/h and a/h definitions of sin and cos.

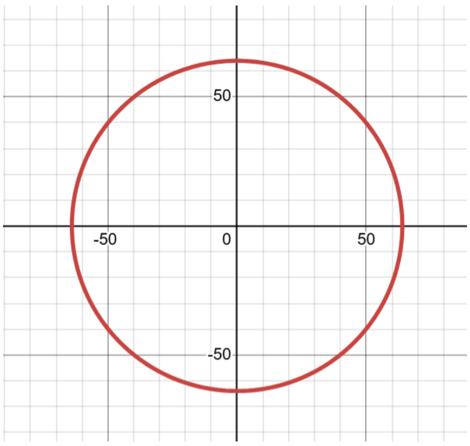


Also, many polynomials have roots in the real numbers, but some, such as $x^2 + 1 = 0$ do not. However, that polynomial has a root in the complex numbers. It turns out that all polynomials with complex coefficients have a root in the complex numbers, and this is called the fundamental theorem of algebra. In the A level documents, we mentioned this, but did not give the proof as we did not need to. However, there is a nice visual argument for why this is the case.

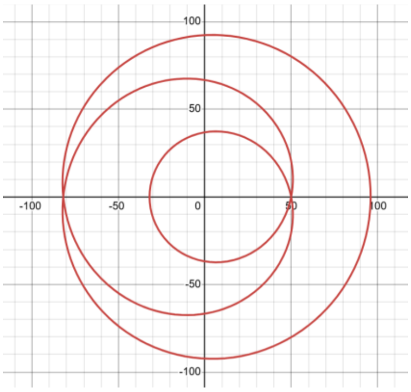
Once I have proven this, we know that by repeatedly factoring out z-(our root) from any polynomial in z until we are left with a constant, that all polynomials are a product of linear factors.

To prove this, I will approach it by first doing an example to show the idea, and then provide the general proof.

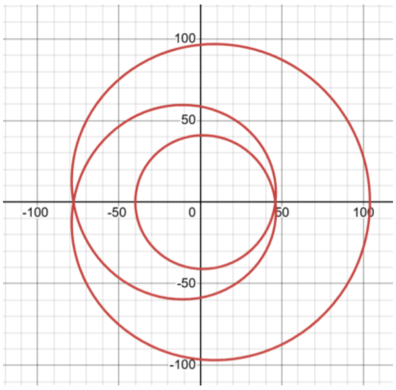
As an example, we will prove that $z^3 + 2z^2 + 2z + 3$ has a root in $\mathbb{C}$. Suppose $z = re^{it}$ where t can go from 0 to 2π and r is picked to be sufficiently large (we will show this can always be done). Suppose r=4 for this example. Now consider z^3 as we let t change. This is just $64e^{3it}$ which is a circle of radius 64.



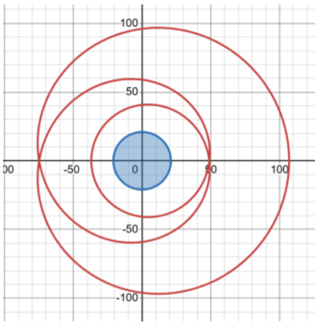
Now lets continuously add $2z^2 = 32e^{2it}$ to each point. If you picture each point moving to its new position, you will see they cannot move more than 32 units, which means you can see that the new loop must enclose the circle centered at the origin of radius 32 as we can never enter this circle during the continuous transformation.



Now lets add $2z = 8e^{it}$. Something not inside a radius 32 circle moving by 8 units surely won't be inside a radius 24 circle. Here's the new diagram, which must enclose the radius 24 circle.



Finally, we add 3 and we are now enclosing the radius 21 circle for similar reasons.



Now continuously shrink the value of r. As this happens, our diagram will continuously shrink to the point at 3. Therefore, at some point during the shrinking, our loop will have to touch the origin. There is no way to continuously move the loop to the point at 3 without breaking it and without touching the origin. When we do touch the origin, **the value of z corresponding to that point is a root of the polynomial, meaning a root exists. This is the core idea.** This happens at the point where we go from enclosing the origin to not enclosing the origin, which exists and is somewhere between 0 and 4. Everything is continuous here because it is differentiable, but we will try to avoid getting stuck on trivial details like these.

So, in general, to pick r large enough, we can make it larger than twice the absolute value of the largest coefficient. This will ensure that the minimum enclosed circle radius which went from 64 to 32 to 24 to

21 in our above example will stay positive, by a geometric series bounding argument (since each difference will be at most half of the previous one so the sum of all the differences will be no larger than the starting value). So I should have picked r=6 in the first example, but 4 turned out to be good enough for our purposes.

Lecture 2:

We reviewed some complex number properties from A levels, including some of the stuff about branches from the exponentials and logarithms video in level 4. However, we do prove two new equations that describe lines and circles.

A line in the complex plane can be described by a point on the line z_0 and a direction ω so that the line is all points $z_0 + a\omega$ with a real. Write $z = z_0 + a\omega$. Taking conjugates on both sides gives $\bar{z} = \bar{z_0} + a\bar{\omega}$. For now we will assume $|\omega| = 1$. Then $\overline{z - z_0} = \bar{z} - \bar{z_0} = a\bar{\omega}$, since recall it is straight forward that $\overline{a \pm b} = \bar{a} \pm \bar{b}$ and that $\overline{ab} = \bar{a}\bar{b}$ as can be verified by splitting everything into its real and imaginary parts and subtracting imaginary parts for the conjugate. Since $\omega^{-1} = \bar{\omega}$ as $|\omega| = 1$ (since $\omega\bar{\omega} = |\omega|^2 = 1$), we therefore have $(\overline{z - z_0})\omega = a$, and substituting this into $z = z_0 + a\omega$ gives $z = z_0 + (\overline{z - z_0})\omega^2$. Dividing both sides by ω gives $z = z_0 + (\overline{z - z_0})\omega$. Rearranging this gives:

$z\bar{\omega} - \bar{z}\omega = z_0\bar{\omega} - \bar{z_0}\omega$ as the general equation for a line. In fact, we no longer require $|\omega| = 1$ as now we can just divide this equation by $|\omega|$ to get an equation of the form where the thing being multiplied by z does have an absolute value of 1.

A circle can be written as $|z - z_0| = r$ and therefore $|z - z_0|^2 = |r|^2$, so $(z - z_0)(\overline{z - z_0}) = |r|^2$, so $|r|^2 = z\bar{z} + z_0\bar{z_0} - z\bar{z_0} - \bar{z}z_0 = |z|^2 + |z_0|^2 - z\bar{z_0} - \bar{z}z_0$. So we can write the equation for a circle in the complex plane as

$|z|^2 - z\bar{z_0} - \bar{z}z_0 = |r|^2 - |z_0|^2$.

Also, recall that any complex number z can be written as $re^{i\theta}$. If $\theta \in (-\pi, \pi]$ then we can write $\log(z) = i\theta + \log(r)$ where log is the natural logarithm. The idea is that arg(z) is the imaginary part of log(z), and Arg(z) which is the multi-valued argument, ie arg(z)+2nπ is the imaginary part of Log, where a capital letter at the start means the multi valued counterpart.

Lecture 3:

A vector can be thought of as an ordered tuple of real or complex numbers, or a line segment, or a thingy with a magnitude and a direction. If a vector $v = \overrightarrow{AB}$ then the direction of v is B-A.

Definition: A **vector space** V (over $\mathbb{R}$ or $\mathbb{C}$) is a set of vectors that satisfies some properties. Examples of vector spaces are $\mathbb{R}^n$ and $\mathbb{C}^n$ where n is some finite number.

Property 1: We can add vectors, and addition satisfies that for vectors a, b in V, a+b is in V, a+b=b+a (commutitivity), (a+b)+c=a+(b+c) (Associativity), and there exists an identity vector 0 such that 0+a=a for all a, and for each a there exists a vector -a such that a+(-a)=0 (Inverses). This is exactly the axioms for an abelian group.

Property 2: We can multiply vectors by scalars, where a scalar is an element of the set which our vector space is over (which is usually the real numbers or the complex numbers). Scalar multiplication satisfies the obvious properties: For any scalars λ and μ, λ(a+b)= λa+λb, (λ+μ)a=λa+μa, λ(μa)=μ(λa)=(μλ)a, and 1a=a.

Example: $\mathbb{R}^n$ is the set of n dimensional vectors which are written as a list with n components that are each real-valued with addition and multiplication defined component-wise (ie add and multiply each component). A line in $\mathbb{R}^n$ is a vector space if and only if it goes through the origin, as that is necessary and sufficient for it to satisfy some of the given properties.

Definition: If I have vectors a and b and scalars λ and μ, then any vector of the form λa+μb is called a **linear combination** of a and b, and analagously if we had more than 2 vectors and corresponding scalars. The set of all such vectors is called the **span** of a and b.

Definition: a is parallel to b if a=λb for some scalar λ or b=0. If a is not parallel to b then the span of a and b is the plane through a, b and the origin, and this is geometrically obvious.

We have met the dot product at A level before. Note that it is only defined for real-valued vectors, for complex-valued vectors it works a bit differently. Here are some obvious properties of dot products that follow immediately from either the geometric or algebraic definition, which we showed were equivalent in A level.

- (λa).b=λ(a.b)
- From the above property and the fact that we showed in A level that a.(b+c)=(a.b)+(a.c), we have that (λa+μb).c=λ(a.c)+μ(b.c)

Note that a.a is the sum of the squares of the components of a by the algebraic definition of the dot product, so it is equal to $|a|^2$ by pythagoras. We can actually define $|a|$ this way as $\sqrt{a.a}$, and $|a|$ is a **norm**. Norms are usually written as either $|a|$ or $||a||$, and they satisfy that |a|=0 if and only if a=0, always being positive, as well as the triangle inequality.

We can get the cosine rule directly from the geometric definition of the dot product: For a triangle ABC, $\overrightarrow{BC} = \overrightarrow{AC - AB}$ so $\left|\overrightarrow{BC}\right|^2 = \left|\overrightarrow{AC - AB}\right|^2 = \overrightarrow{AC - AB}.\overrightarrow{AC - AB} = \overrightarrow{AC}.\overrightarrow{AC} + \overrightarrow{AB}.\overrightarrow{AB} - 2\overrightarrow{AC}.\overrightarrow{AB} = \left|\overrightarrow{AC}\right|^2 + \left|\overrightarrow{AB}\right|^2 - 2\left|\overrightarrow{AC}\right|\left|\overrightarrow{AB}\right|\cos(BAC)$ by the geometric definition of the dot product.

Lecture 4:

The Cauchy-Schwartz inequality says $|a.b| \leq |a||b|$, which is normally obvious in vector spaces $\mathbb{R}^n$ since $|cosx| \leq 1$ for real x. However, we can prove this for more general vector spaces:

We know $|x - ay|^2 \geq 0$ for all a since the norm of a vector is always non-negative by definition. Therefore $(x - ay).(x - ay) \geq 0$, so $(x.x) + a^2(y.y) - 2a(x.y) \geq 0$. Therefore $|x|^2 + a^2|y|^2 - 2a(x.y) \geq 0$. This is a quadratic in a, and for it to always be positive it cannot have 2 roots so the discriminant must be non-positive, therefore $4(x.y)^2 - 4|x|^2|y|^2 \leq 0$. Rearranging and taking the square root of both sides gives $|a.b| \leq |a||b|$ as required.

Also we have the triangle inequality: $|x + y|^2 = (x + y).(x + y) = |x|^2 + |y|^2 + 2(x.y) \leq (|x| + |y|)^2$ by the previous inequality, thus $|x + y| \leq |x| + |y|$.

Definition: Vectors $v_1, v_2, v_3, \ldots$ are **linearly independent** if $a_1v_1 + a_2v_2 + \cdots a_nv_n = 0$ implies all the a's are equal to 0. Ie, no vectors are a linear combination of the others.

Definition: A **basis** for a vector space is a set of linearly independent vectors that span the vector space. A basis is often written as $\{e_1, e_2, \ldots\}$ in general or {i, j, k} in the case of $\mathbb{R}^3$.

Note that the cross product is a thing specific to $\mathbb{R}^3$.

Recall some easy properties:

$a \times b = 0$ if and only if a, b are parallel.

$c(a \times b) = (ca \times b) = (a \times cb)$

$a \times (b + c) = a \times b + a \times c$

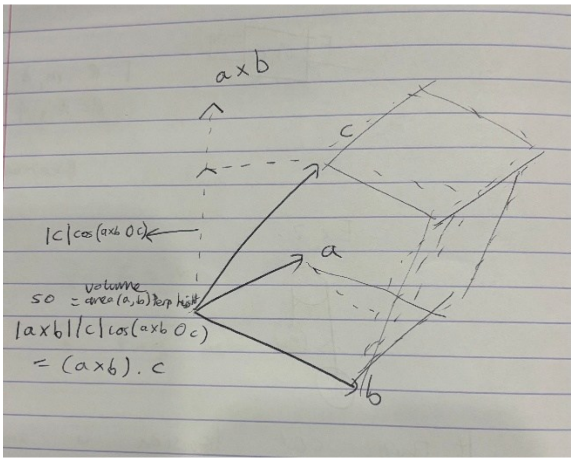
$a \times b = -b \times a$

These can be shown by determinant properties as the cross product was interpreted as.

$$\begin{vmatrix} i & a_i & b_i \\ j & a_j & b_j \\ k & a_k & b_k \end{vmatrix} = a \times b.$$

Also, $a.(a \times b)$ since $a \times b$ is perpendicular to a, and because in fact $(a \times b).c$ gives the volume of the parallelapiped spanned by a, b, c, and thus is zero exactly when a, b, c are coplanar. This is called the scalar triple product

To see this, here is a little picture, note that the volume of the parallelopiped is the area spanned by a and b times the component of c in the direction of axb.



Lecture 5:

Interpretation: A×X is X scaled by |a| and rotated 90 degrees in such a way as to be perpendicular to a.

The lecturer also goes from $(a_1e_1 + a_2e_2 + a_3e_3) \times (b_1e_1 + b_2e_2 + b_3e_3)$ and uses the basic cross product properties above to reach the algebraic formula for the cross product.

The vector triple product is defined as a×(b×c) and gives a vector perpendicular to a in the same plane as b and c. By expanding definitions it can be shown that a×(b×c)=b(a.c)-c(a.b), but we will give a nicer proof of this later. Note that a×(b×c) is not generally equal to (a×b)×c so we need to be careful.

Recall that a line through a in the direction of u can be written as (r-a)×u=0 and that a plane through a with direction vectors u and v can be written as r.n=a.n with n=(u×v) and thus (r-a).(u×v)=0.

Therefore, we can give a general formula for the intersection between a line and a plane by solving for r:

(r-a)×u=0 by the line equation

So ((r-a)×u)×n=0 by taking the cross product with n on both sides. Using antisymmetry twice so the sign cancels we get Using the formula for the vector triple product, we get that n×(u×(r-a))=0

u(n.(r-a))-(r-a)(n.u)=0.

Therefore, u(n.r)-r(n.u)=u(n.a)-a(n.u)=(a×u)×n.

Let b be any point on the plane, then u(n.b)-r(n.u)=(a×u)×n so $r = \frac{(b.n)u-(\mathrm{a}\times\mathrm{u})\times\mathrm{n}}{u.n}$.

Lecture 6:

Lets also consider the shortest distance between 2 lines: Suppose the lines go through points a_1, a_2 and have direction vectors u_1, u_2. Then the equation for the lines is $u_1 \times (r - a_1) = 0$ and $u_2 \times (r - a_2) = 0$. The shortest distance is in the direction perpendicular to both lines, so if they are not parallel this is the projection in the direction of $u_1 \times u_2$ of $a_1 - a_2$. Thus the shortest distance between the lines is given by $\left|(a_1 - a_2).\frac{(u_1 \times u_2)}{|u_1 \times u_2|}\right|$.

Here is an example of solving for a vector:

Say we want to solve for r where $r + a \times (b \times r) = c$. Then using the formula for the vector triple product,

$$r + (a.r)b - (a.b)r = c$$

$$r.a + (a.r)b.a - (a.b)r.a = c.a$$

$$r.a = c.a$$

Which is a plane through c perpendicular to a. We can substitute this into $r + (a.r)b - (a.b)r = c$ to get $r + (a.c)b - (a.b)r = c$ so $r = \frac{c-(a.c)b}{1-(a.b)}$. If a.b=1 then either $c - (a.c)b$ is non zero so r is nothing or $c - (a.c)b$ is zero meaning r is the full plane through c perpendicular to a.

New definitions:

An ordered set of vectors a, b, c is right handed if the scalar triple product is positive and left handed otherwise.

Definition (kronecker delta) $\delta_{ij} = \begin{cases} 1: i = j \\ 0: i \neq j \end{cases}$

Definition (Levi civite symbol) $\varepsilon_{ijk} = \begin{cases} 1: i,j,k \text{ } is \text{ } an \text{ } even \text{ } permutation \\ 0: i = j \text{ } or \text{ } j = k \text{ } or \text{ } k = i \\ -1: i,j,k \text{ } is \text{ } an \text{ } odd \text{ } permutation \end{cases}$

We have discussed what even and odd permutations are in groups.

There is a convention with terms which are products involving these symbols. The convention (Einstein's convention) is that if an index like i, j or k appears exactly twice in the expression we sum over the implied values – typically 1, 2 and 3. For example, if $\{e_1, e_2, e_3\}$ is our basis, then $e_i\delta_{ij}$ has i exactly twice so it equals $e_1\delta_{1j} + e_2\delta_{2j} + e_3\delta_{3j}$. The only term that does not vanish is the term where i=j by the definition of the delta, and therefore $e_i\delta_{ij} = \delta_j$. If an index appears more than twice, the convention is not used, but it is surprisingly useful to do it this way.

Examples (I will not prove these since they are easy but tedious to check):

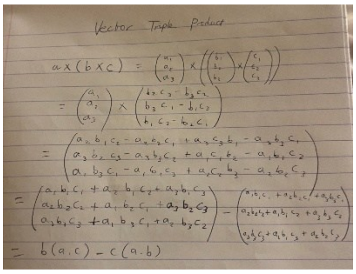
- $e_i . e_j = \delta_{ij}$
- $a. b = \delta_{ij} a_i b_j$
- $e_i \times e_j = \varepsilon_{ijk} e_k$
- $a \times b = a_i b_j e_k \varepsilon_{ijk}$
- $(a \times b)_i = a_j b_k \varepsilon_{ijk}$ (Note: this means the component of $a \times b$ in the direction of the basis vector i)
- $a. (b \times c) = a_i b_j c_k \varepsilon_{ijk}$
- $\delta_{ii} = 3$
- $\varepsilon_{ijk}\varepsilon_{pqr} = \delta_{ip}\delta_{jq}\delta_{kr} + \delta_{iq}\delta_{jr}\delta_{kp} + \delta_{ir}\delta_{jp}\delta_{kq} - \delta_{ip}\delta_{jr}\delta_{kq} - \delta_{iq}\delta_{jp}\delta_{kr} - \delta_{ir}\delta_{jq}\delta_{kp}$
- $\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$
- $\varepsilon_{ijk}\varepsilon_{pjk} = 2\delta_{ip}$
- $\varepsilon_{ijk}\varepsilon_{ijk} = 6$

We will use this to show that a×(b×c)=b(a.c)-c(a.b). Consider the component of a×(b×c) in the direction of a basis vector such as i, then (using the identites above and doing algebra) we get

$$[a \times (\mathrm{b} \times \mathrm{c})]_{\mathrm{i}} = \varepsilon_{ijk}[a_j(b \times \mathrm{c})_{\mathrm{k}}] = \varepsilon_{ijk}[a_j b_p c_q \varepsilon_{pqk}] = (\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp})[a_j b_p c_q]$$

Note that j looks like it appears 3 times but it only appears 2 times in each term when we expand it out so it is fine.

$= \delta_{ip}\delta_{jq}a_j b_p c_q - \delta_{iq}\delta_{jp}a_j b_p c_q = a_j b_i c_j - a_j b_j c_i = (a. c)b_i - (a. b)c_i$ so done. This is much more compact then if we were to expand the vectors out, here is an image to show what that looked like:



Lecture 7:

Another example of using the summation convention to prove an identity:

$$(a \times \mathrm{b}). (\mathrm{b} \times \mathrm{c}) = (\mathrm{a} \times \mathrm{b})_{\mathrm{i}}(b \times \mathrm{c})_{\mathrm{i}} = \varepsilon_{ijk}a_j b_k \varepsilon_{ipq} b_p c_q = (\delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp})a_j b_k b_p c_q$$
$$= a_j b_k b_j c_k - a_j b_k b_k c_j = (a. b)(b. c) - (a. c)(b. b)$$

Now suppose we have a unit sphere with points a, b and c on the sphere. Then $a. b = \cos(\delta(a, b))$ where $\delta(a, b)$ means the arc length from a to b.

Now note that $\frac{a \times \mathrm{b}}{|\mathrm{a} \times \mathrm{b}|}$ and $\frac{a \times \mathrm{c}}{|\mathrm{a} \times \mathrm{c}|}$ are unit vectors perpendicular to the planes through AOB and AOC respectively, so the cosine of the angle between them equals the cosine of the angle between the arcs AB and AC. However, $|\mathrm{a} \times \mathrm{b}| = |a||b| \sin(AOB) = \sin(\delta(a, b))$. However, the cosine between two unit vectors is the dot product, so if α is the angle between the arcs then $\cos(\alpha) = \frac{(a \times \mathrm{b}).(\mathrm{a} \times \mathrm{c})}{\sin(\delta(a,b))\sin(\delta(a,c))} =$

$\frac{|a|^2(b.c)-(a.c)(b.a)}{\sin(\delta(a,b))\sin(\delta(a,c))}$ by the above identity. However, $a.b = \cos(\delta(a,b))$ and similarly for a.c and b.c so $\cos(\alpha) = \frac{\cos(\delta(b,c))-\cos(\delta(a,b))\cos(\delta(b,c))}{\sin(\delta(a,b))\sin(\delta(a,c))}$.

Definition: $\varepsilon_{ijk\ldots l}$ if the indices range from 1 to n where n is the number of indices is 0 if any 2 are the same, 1 if it is an even permutation and -1 if it is an odd permutation. See Lecture 13 for what this means and why it is well defined. This is the generalized epsilon.

Example: The determinant of a 2x2 matrix $\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$ is $\varepsilon_{ij}a_ib_j$. In fact, higher order determinants use the same general formula.

Lecture 8:

Definition (Inner product): This is basically the dot product. (a,b) is an inner product if it satisfies:

- (a,a) is non-negative and always real
- (a,b) is always the complex conjugate of (b,a) (Note that this means that in the real number case we have symmetry)
- $(z, cw + c'w') = c(z,w) + c'(z,w')$
- $(cw + c'w', z) = \overline{c(w,z) + c'(w',z)}$

Definition: A basis is orthonormal if $e_i.e_j = \delta_{ij}$ – ie each basis vector has magnitude 1 and is perpendicular to the rest.

Defintion: A real vector space has multiplication only by real numbers. A complex vector space has multiplication by real or complex numbers. The real vector space defined by $\mathbb{C}^n$ actually has dimension 2n because we need the basis (1,0,0...), (i,0,0,...), (0,1,0,...), (0,i,0,...), etc. We will prove soon that dimension is actually unique.

Definition: A subspace is a subset of a vector space that is also a vector space. For a space to be a subspace, we need that for any vectors v and w their span (ie all linear combinations) is in the subspace. V itself and {0} are trivial subspaces of any vector space V.

More on linearly independent vectors:

Example: (3, -1, 2), (1, 0, 1) and (5, -2, 3) are not linearly independent because (5, -2, 3)=2(3, -1, 2)-(1, 0, 1). However the set (3,0,0), (4,1,3) and (2,0,1) are linearly independent. This can be checked: $\begin{vmatrix} 3 & 4 & 2 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{vmatrix} = 3$ so the vectors span three dimensions (geometrically). Note that vectors being pairwise linearly independent is not enough. However (geometrically), mutually perpendicular vectors in real vector spaces are linearly independent.

Proof: Let $(v_1, v_2, \ldots, v_n)$ be mutually perpendicular and non-zero. Then $(v_j, \sum a_i v_i) = \sum a_i\,(v_j, v_i)$ (but unless i=j, the terms are 0), so we get $= \sum a_j\,(v_j, v_j) = \sum a_j\,|v_j|^2$, so if the sum is 0 all a's must be 0 since they are being multiplied by positive things, so the v's satisfy the definition of linearly independent.

Defintion: The dimension of a vector space is the number of basis vectors that span the vector space. However, we need to show that this is the same regardless of which basis we use.

Proof: (Credit: Google images – Image below shows proof)

Claim: If $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\}$ is linearly independent in V and $\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \ldots, \vec{w}_m\}$ spans V, then $n \leq m$.

Strategy: Replace the vectors of our spanning set with those of our linearly independent set, one by one.

Proof: $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\}$ is linearly independent in V.

$$\vec{v}_1 = \alpha_1\vec{w}_1 + \alpha_2\vec{w}_2 + \alpha_3\vec{w}_3 + \cdots + \alpha_m\vec{w}_m$$

without loss of generality, $\alpha_1 \neq 0$; replace $\vec{w}_1$ with $\vec{v}_1$ to get a new spanning set.

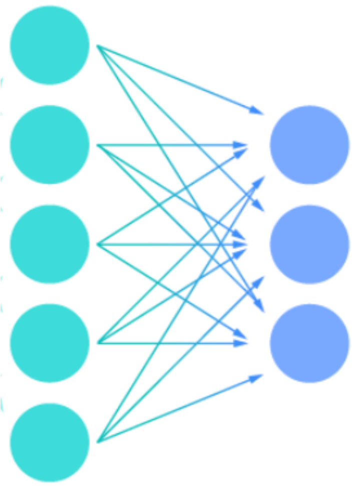
$\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \ldots, \vec{w}_m\}$ spans V.

Therefore, suppose we have a basis with X vectors and another one with Y vectors for the same vector space, and suppose X<Y (since if Y<X we could apply similar logic without loss of generality): Then the Y basis is linearly independent and the X basis spans the vector space, but X<Y so this contradicts the theorem above.

Of course, if we have a set that spans a vector space we can remove redundant ones to get a basis. Also, if we have a set that does not span a vector space, we can add new ones until it does then remove redundant ones.

Lecture 9:

Consider a map from a bunch of numbers to a bunch of other numbers. Sort of like this image below from google images where each arrow has a weight and the output is the weighted sum.



This is called a linear map, and you can see that this is equivelent to multiplying the left hand side as a vector on the left by a matrix with entries equal to the weights of the arrows in the corresponding positions. In fact, matrices of a fixed size are elements of a vector space since they satisfy all the elements of a vector space. You can think of a matrix as a map from one vector space to anoyher vector space.

Definition: The **image** of a matrix is the set of vectors that get mapped to. The **rank** of a matrix is the dimension of the image. The **kernel** of a matrix is the set of vectors in the domain that map to zero. The **nullity** or null space of a matrix is the dimension of the kernel. Kernels and images are closed under adding or subtracting them and stuff and contain the origin so they do form vector spaces – this is important to know.

Theorem (Rank-Nullity theorem): If we have an nxn matrix, then its rank plus its nullity equals n.

Proof: This is a direct consequence of the isomorphism theorem which we learn in the Groups course. This is because if we consider our matrix to be a homomorphism between our vector spaces as groups under addition, then the vector space we are mapping to is isomorphic to the direct product of the kernel and the image, and therefore the sum of the dimensions of the kernel and the image is n.

Note: This is obvious if you consider volume, but a square matrix maps to the entire vector space if and only if its determinant is non-zero. Also, since such a matrix is invertible we can use its inverse to find the pre-image of any vector.

Note: A matrix can be interpreted as a list of vectors where each row represents a vector, or where each column represents a vector. Ie

$$\begin{matrix} \vdots & \vdots & \cdots & \vdots \\ C_1 & C_2 & \cdots & C_n \\ \vdots & \vdots & \cdots & \vdots \end{matrix}$$

And similarly for rows.

Note that the span of the columns of a matrix is the column space or span of a matrix. This is also equivelent to the image of the matrix, since the image is exactly linear combinations of the columns with coefficients equal to the entries we are multiplying the matrix by.

Lecture 10:

We note that multiplying a matrix by a column vector like (0,0,…,0,1,0,…,0,0) essentially extracts a column from the matrix.

We note also that $Mx = \begin{pmatrix} R_1.x \\ R_2.x \\ \vdots \\ R_n.x \end{pmatrix}$. This is easy to check, where the R's are the rows of M.

Therefore, it follows that $Mx = 0$ when x is a vector perpendicular to all rows of the matrix.

Example: The rank of $\begin{pmatrix} 3 & 1 & 5 \\ -1 & 0 & -2 \\ 2 & 1 & 3 \end{pmatrix}$ is 2 because

- It is not 0 since 0 is clearly not the image
- It is not 1 since all columns do not lie on the same line
- It is not 3 since the determinant is 0.

Therefore, by rank nullity, its kernel must be a line.

Definition: In 2D, $Rot(\theta)$ is a matrix representing a rotation by an angle θ anticlockwise. In fact, by checking where (0,1) and (1,0) go we see that $Rot(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. This is known from A level. Similarly, we define $Ref(\theta)$ as the matrix representing a reflection about the line which is an angle $\frac{\theta}{2}$ anticlockwise from (1, 0). If you check carefully where basis vectors go, we see that $Rot(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}$. We can try multiplying them together in different ways:

- $Rot(\theta)Rot(\phi) = Rot(\theta + \phi)$. I think this is obvious if you think about it.
- $Ref(\theta)Ref(\phi) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}\begin{pmatrix} \cos(\phi) & \sin(\phi) \\ \sin(\phi) & -\cos(\phi) \end{pmatrix} = \begin{pmatrix} \cos(\theta-\phi) & -\sin(\theta-\phi) \\ \sin(\theta-\phi) & \cos(\theta-\phi) \end{pmatrix} = Rot(\theta - \phi)$ where I have used trigonometric addition formulae which are known from A level.
- $Ref(\theta)Rot(\phi) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}\begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} = \begin{pmatrix} \cos(\theta-\phi) & \sin(\theta-\phi) \\ \sin(\theta-\phi) & -\cos(\theta-\phi) \end{pmatrix} = Ref(\theta - \phi)$
- $Rot(\theta)Ref(\phi) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}\begin{pmatrix} \cos(\phi) & \sin(\phi) \\ \sin(\phi) & -\cos(\phi) \end{pmatrix} = \begin{pmatrix} \cos(\theta+\phi) & \sin(\theta+\phi) \\ \sin(\theta+\phi) & -\cos(\theta+\phi) \end{pmatrix} = Ref(\theta + \phi)$.

In 3 dimensions, a rotation of an angle θ anticlockwise about a unit vector n is given by a formula which we will derive next lecture. The formula says that a vector x is mapped to

$\cos(\theta)\, x + (1 - \cos(\theta))(n.x)n + (\sin(\theta))(n \times x)$.

A matrix representing a reflection around a space A is given by $I - 2AA^T$ where A consists of columns of normalized vectors that are perpendicular and span A. The reason is that the formula for an orthogonal projection onto A is known to be $I - AA^T$ as I show shortly, and it should make sense that an orthogonal projection moves a vector halfway from its original position to its reflected position.

We are in $\mathbb{R}^n$ (n dimensional vector space of reals) and for m<n there is an m-dimensional plane centered at the origin, then we can rotate the space around so that the first m basis vectors are in our plane, then consider where the rest of the basis vectors were before we did the rotation: Intuitively, we are considering a basis for vectors perpendicular to this plane. If S is a n*(n-m) matrix where each column of s is one of these vectors, we want to show that $I - SS^T$ projects any vector onto the m-dimensional plane. To do this, we just need to show three things:

1. A vector when this linear transformation is applied to it moves orthogonally to the direction of the plane
2. Any vector ends up on the plane after the transformation

Consider a vector v, then $(I - SS^T)v = Iv - SS^Tv = v - SS^Tv$

Condition 1: Orthogonality

The vector moves by SS^Tv. Therefore we want to show that $(SS^Tv).u = 0$ if u is on the plane. This is the same as $(SS^Tv)^Tu = v^TSS^Tu$. But we know S^Tu is 0 since we are assuming u is on the plane (so all columns of S as a dot product with v return 0 so the result follows), so the whole thing becomes 0. So done.

Condition 2: A vector ends up on the plane after the transformation

We want to show $(I - SS^T)v$ ends up on the plane. Since the plane is defined by S^Tu is 0 if u is on the plane, we want to show that $S^T(I - SS^T)v = 0$. This is equal to $S^Tv - S^TSS^Tv = (I - S^TS)S^Tv$.

Now what does $S^T S$ equal? It will be an (n-m)x(n-m) matrix where the i,j entry is equal to $s_i . s_j$. Since the s's are orthogonal unit vectors, this will be 1 when i=j and 0 otherwise, so we get the identity matrix. Therefore $I - S^T S = 0$. So done.

Now suppose B is a matrix where the columns are vectors that form an orthonormal basis of the plane (Essentially this means what you would expect: Where the basis vectors on the plane were before the rotation). Then $I - BB^T$ is the projection matrix onto the space perpendicular to the plane, since it is essentially the same idea with S and B being renamed to eachother, as they are both matricies with perpendicular columns within them and between them. Now a vector v decomposes into vectors v_s and v_b, where v_b is the component of v in the direction of the plane, and v_s is the component of v perpendicular to the plane. So $(I - BB^T)v = v_s$ and $(I - SS^T)v = v_b$, so $(I - BB^T)v + (I - SS^T)v = v_s + v_b = v$. So$(2I - BB^T - SS^T)v = v$. Since this is true for all v, we must have that $I = 2I - BB^T - SS^T$ so $I = BB^T + SS^T$. What I have proven is that if B and S are matricies whose columns together form a basis for a vector space with all vectors perpendicular to eachother then $I = BB^T + SS^T$. Alternatively, I have shown that if B forms a basis for the plane with all vectors in B perpendicular to eachother, then BB^T gives a projection onto the plane.

Lecture 11:

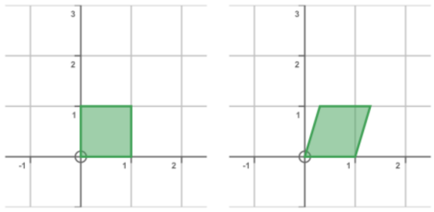
Now we will derive the rotation formula as promised. We can write $x = x_n + x_p$ where x_n is the component of x parallel to n and x_p is the component of x perpendicular to n. Since |n|=1, we know that $x_n = |x| \cos(nOx)\, n = (n.x)n$. Also $x_p = x - x_n$. We know that when x is rotated about n, the x_n component will not change at all since anything parallel to n does not move when rotated about n. Now we must work out what happens to x_p. It is rotated by θ about n, so it must have a part equal to $\cos(\theta)\, x_p$ and a part equal to $\sin(\theta)\, y$ where y is an anticlockwise rotation of x about n. Now consider n×x: By the right (left? I'm not sure it doesn't matter) hand rule, this is perpendicular to n and a 90 degree rotation anticlockwise from x, and its magnitude is exactly |n||x|sin(nOx) which is |x| since the other terms in the product are 1. This means n×x is exactly the y vector which we need. Therefore, x is mapped to $x_n + x_p\, cos(\theta) + sin(\theta)\,(n \times x) = (n.x)n + (x - (n.x)n)\, cos(\theta) + sin\,(\theta)(n \times x)$. Rearranging gives the desitred formula.

In fact, it turns out that any rotation in 3 dimensions that moves the 3 basis vectors to any other orthonormal right handed set can be written as the product of 3 rotations about the axes. Here is why:

Here is how: We rotate about the x axis first to move the z axis vector to have the correct final height. We then rotate about the original z axis to make this vector be in the correct final position. The x and y vectors may be wrong, but we can pre-rotate them about the z axis before doing these other 2 rotations such that they end up in the correct place.

Definition: Suppose a and b are perpendicular and |a|=|b|=1. Then a shear is a matrix which sends x to x+λa(x.b). Basically, this leaves a unchanged and moves b by λa units.

This image from google below shows an example of a shear where we leave the x axis unchanged and slightly shift the y axis over.



Also, here is an example of the principle of a matrix being a linear map between vector spaces: Consider a map from the vector space of 2x2 matrices to 3x1 vectors (ie, $\mathbb{R}^3$). Suppose the map always sends $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \to \begin{pmatrix} a+b \\ c \\ d \end{pmatrix}$. Then we have to write our matrix as a column vector: $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$. Then the matrix that sends this to $\begin{pmatrix} a+b \\ c \\ d \end{pmatrix}$ is (This can be easily checked, since a map is defined in the matrix multiplication sense) $\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$.

An important idea is that this map is really matrix multiplication is the same as composing maps. This is because if A: V -> W and B: W->S then AB is the map (A∘B) from V to S, where V, W and S are vector spaces. This works like this due to matrix multiplication being associative: (AB)X=A(BX) if X is a vector in our starting vector space V.

Lecture 12:

Matrix multiplication can be interpreted by the summation convention as follows: if L=MN then $L_{ik} = M_{ij}N_{jk}$. Also, we can see that L_{ik} is the dot product of the i'th row of M with the k'th column of N. In fact, associativity of multiplication follows directly from this: $A_{ij}(B_{jk}C_{kl}) = (A_{ij}B_{jk})C_{kl}$.

Note that a left inverse of a square matrix is the same as a right inverse with the same proof as at the beginning of groups. We were able to see that this was the case in the Level 6 matrix video.

Some obvious properties: An inverse of a rotation is a rotation in minus the angle. Reflections are self inverse. An inverse of a shear is a shear the opposite way.

Definition: The **hermitian conjugate** of a matrix is the complex conjugate of its transpose. This is denoted $M^{\dagger}$.

Definition: A matrix is **antisymmetric** if $M^T = -M$. This actually constrains the matrix a lot – All diagonal entries in an antisymmetric matrix have to be 0.

Definition: The **trace** of a square matrix is the sum of the elements along the diagonal. By the summation convention we can write $Tr(M) = M_{ii}$. By considering the characteristic equation we can see that the trace equals the sum of the eigenvalues.

Proposition: $Tr(MN) = Tr(NM)$. Proof: $(MN)_{ii} = M_{ik}N_{ki} = N_{ki}M_{ik} = (NM)_{kk}$ (summation convention used).

Proposition: Any matrix with real entries is a sum of antisymmetric parts.

Proof: Set $S = \frac{1}{2}(M + M^T)$, $A = \frac{1}{2}(M - M^T)$. It is easy to see that S is symmetric, A is antisymmetric, and S+A=M.

S can be further decomposed into $T \coloneqq S - \frac{1}{n}Tr(S)I$ and $\frac{1}{n}Tr(S)I$. Note that the trace of T is 0 by how we constructed T.

Therefore, M has been decomposed into a symmetric traceless matrix, an antisymmetric matrix and a multiple of the identity matrix.

Lecture 13:

We have seen in Level 6 in the section on symmetric matrices a dot product argument for why a matrix U is orthogonal if and only if $UU^T = U^TU = I$, ie the columns and rows of U are orthonormal. Orthogonal matrices are defined to be only real-valued matrices with this property.

Also, $(Ux).(Uy) = x.y$ if U is orthogonal: This is because if U is orthogonal it is essentially a rotation, meaning the lengths and angles – and thus the dot product – are unchanged.

Definition: A complex n*n matrix is unitary if its inverse equals its hermitian conjugate. This is a generalization of orthogonal matrices. If U is unitary, we have, using the complex version of the dot product, that $(Uz).(Uw) = (Uz)^\dagger(Uw) = z^\dagger U^\dagger Uw = z^\dagger w = z.w$.

Example: If $U\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}$, then $U\begin{pmatrix}0\\1\end{pmatrix} = \pm\begin{pmatrix}-\sin(\theta)\\ \cos(\theta)\end{pmatrix}$. Basically, angles and lengths must be preserved for unitary matrices, but not orientation.

Example: The determinant of a 3*3 matrix M can be written in index notation as $\varepsilon_{ijk}M_{1i}M_{2j}M_{3k}$. For higher size matrices, the determinant can indeed be written as $\varepsilon_{ijk\dots l}M_{1i}M_{2j}M_{3k} \dots M_{nl}$ using the generalized epsilon. The fact that the three dimensional case of the determinant is the volume was proven in Level 6, and the higher dimensional case uses the same proof provided we have well defined-ness of the sign of the permutation as a theorem. We will prove this later in the lecture.

It is easy to see that matrix columns are all linearly independent if and only if the determinant is non-zero.

Proof: If the determinant is zero, then everything is confined to a lower dimension by the volume property, and therefore they cannot be spanned by n linearly independent vectors. Conversely, if the matrix has columns all linearly independent then it has full rank (so it is surjective) and thus trivial kernel, and therefore it is injective (Since if two vectors mapped to the same thing under the matrix then the difference would be in the kernel and thus would be 0). Therefore, if the matrix has linearly independent columns, it is a bijection, and thus is invertible, and thus has non-zero determinant, since if it had zero determinant it would not be invertible as that would involve dividing by 0 which is a contradiction.

Notation for permutations:

Consider the permutation $\begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6\\ 5 & 6 & 3 & 1 & 4 & 2\end{pmatrix}$ where the bottom row shows what the elements of the top row get mapped to. Then let's trace it:

1 maps to 5 which maps to 4 which maps to 1 so we're back to where we started.

2 maps to 6 which maps to 2 so we're back to where we started.

3 maps to itself. Now we have considered all the elements.

So, a convention is we can write $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 3 & 1 & 4 & 2 \end{pmatrix}$ as $(1\ 5\ 4)(2\ 6)$, ie as a list of the cycles described above with the fixed points (like 3) omitted.

Note that we can get the cycle $(1\ 5\ 4)$ by swapping 4 with 5 then swapping 5 with 1. In general, we can get a cycle $(a_1\ a_2 \dots a_n)$ by swapping a_n with a_{n-1}, then a_{n-1} with a_{n-2}, and so on until we swap a_2 with a_1. You can try to trace in your head where anything would get mapped to in order to convince yourself of this.

Definition: The sign of a permutation is 1 if it is the product of an even number of swaps and -1 if it is the product of an odd number of swaps. This is equal to the generalized epsilon. We need to show that this is well defined: We cannot get to a permutation in an even number of swaps and get to the same permutation in an odd number of swaps.

Proof: The idea is a formula exists based on a permutation which changes sign every time we swap two elements, meaning the parity (ie odd-or-even-ness) of how many swaps we did is fixed. We can define the Vandermonde polynomial as $P(x_1, x_2, \dots x_n) = \prod_{1 \le i < j \le n}(x_i - x_j)$. So, for example, $P(1,2,3) = (1-2)(1-3)(2-3) = -2$. Now consider that happens when we swap two elements: $P(x_1, x_2, \dots, x_r, \dots, x_s, \dots, x_n)$ has a certain value, and $P(x_1, x_2, \dots, x_s, \dots, x_r, \dots, x_n)$ has all terms in the product not involving x_r and x_s unchanged. Lets look at what the product of terms involving r and s are in the first and second polynomial.

FIRST ONE:

$$(x_r - x_1)(x_r - x_2)\dots(x_r - x_{r-1})(x_{r+1} - x_r)\dots(x_s - x_r)\dots(x_n - x_r)(x_s - x_1)(x_s - x_2)\dots(x_s - x_{r-1})(x_s - x_{r+1})\dots(x_s - x_{s-1})(x_{s+1} - x_s)\dots(x_n - x_s)$$

Where I have been careful to not include the $x_s - x_r$ term twice.

SECOND ONE:

$$(x_s - x_1)(x_s - x_2)\dots(x_s - x_{r-1})(x_{r+1} - x_s)\dots(x_{s-1} - x_s)(x_r - x_s)(x_{s+1} - x_s)\dots(x_n - x_s)(x_r - x_1)(x_r - x_2)\dots(x_r - x_{r-1})(x_r - x_{r+1})\dots(x_r - x_{s-1})(x_{s+1} - x_r)\dots(x_n - x_r)$$

Now let's cancel terms that are shared to try to see what the ratio is:

The ratio is $\frac{(x_{r+1}-x_r)\dots(x_s-x_r)(x_s-x_{r+1})\dots(x_s-x_{s-1})}{(x_{r+1}-x_s)\dots(x_{s-1}-x_s)(x_r-x_s)(x_r-x_{r+1})\dots(x_r-x_{s-1})}$. Let's write this as follows:

$\frac{(x_{r+1}-x_r)\dots(x_{s-1}-x_r)(x_s-x_r)(x_s-x_{r+1})\dots(x_s-x_{s-1})}{(x_r-x_{r+1})\dots(x_r-x_{s-1})(x_r-x_s)(x_{r+1}-x_s)\dots(x_{s-1}-x_s)} = (-1)^{1+2(s-1-r)} = -1$. So if f is a permutation of (1, 2, 3, … ,n), we can define sign(f) as $\frac{P(f(1), f(2), f(3), \dots, f(n))}{P(1,2,3,\dots,n-1,n)}$, and this is the same as the definition of the sign given above. Therefore this is well defined, so done.

Lecture 14:

Definition: A function F of multiple vectors is **multilinear** if when you fix all but one vectors and treat F as a function of the remaining vector, $F(aV + bW) = aF(V) + bF(W)$ for all a,b,V,W. An example of this is the determinant as a function of the columns of a matrix – We demonstrated why this is the case in the Level 6 matrix video. Multilinear means the same thing even if these are not vectors. Bilinear means multilinear in the case there are two arguments to the function.

F is totally antisymmetric if swapping two arguments changes the sign (ie multiplies it by -1). We are essentially writing names for the determinant properties I showed in that video.

Alternative proof that linear dependence implies zero determinant which is more algebraic. We are doing this because one of the aims of this course is to have you thinking about the connection between the algebraic and geometric side of things:

If the columns are linearly dependent, some column is a linear combination of the others. So we can write $C_p = \sum C_i \lambda_i$. Since subtracting one column from another does not change the determinant, we can subtract λ_i times all the non-p columns from the p'th column to get the p'th column to be 0. So the determinant is 0.

Proposition: For an n*n matrix M, $\det(aM) = a^n \det(M)$. The reason is because if we multiply each of the columns by a, we multiply the determinant of M by a n times, and thus by a^n.

Proposition: The detrminant of a matrix equals the determinant of its transpose. This is because if i,j,k,...l is a permutation, we can take $\det(M) = \varepsilon_{ijk\ldots l} M_{1i} M_{2j} M_{3k} \ldots M_{nl}$. Then $\det(M^T) = \varepsilon_{ijk\ldots l} M_{i1} M_{j2} M_{k3} \ldots M_{ln}$ by definition of the transpose. Now we can apply the inverse of the permutation i,j,k,...l to all the matrix indices in $\varepsilon_{ijk\ldots l} M_{i1} M_{j2} M_{k3} \ldots M_{ln}$ since we will sum over the same terms. that is just reordering the terms, then this will give us $\varepsilon_{i'j'k'\ldots l'} M_{1i'} M_{2j'} M_{3k'} \ldots M_{nl'}$. The sign of the inverse permutation that maps 1, 2, 3, ... ,n to i', j', k', ... l' is the same as the sign of the starting permutation which is why I can prime everything in the epsilon symbol. But summing over i', j', k', ..., l' gives the same result as summing over i,j,k, ..., l. So done.

Remark: Now everything we did for columns of matrices hold equally for rows. It follows that it is the case that the rows of a square matrix are linearly independent if and only if the columns are by considering the determinant. And adding a multiple of a row of a matrix to another row does not change the determinant.

Proposition: Det(MN)=Det(M)Det(N). This is obvious by the volume property, but next lecture we will give an algebraic proof.

Geometrically, if M is orthogonal, its determinant is 1 or -1. Also, since its inverse equals its transpose, the determinant of its transpose must equal the reciprocal of the determinant, so the determinant equals its own reciprocal so it must be -1 or 1.

If M is unitary, then its inverse is the complex conjugate of the transpose of M, so the determinant is the complex conjugate of the determinant of M. Multiplying this by the determinant of M must give 1 since M times its hermitian cnojugate is the identity, so det(M) has modulus 1.

Lecture 15:

Note that swapping columns of a matrix n times multiplies the determinant by $(-1)^n$.

$$Det(MN) = \sum_{\sigma \in S_n} \varepsilon(\sigma) (MN)_{\sigma(1)1} (MN)_{\sigma(2)2} (MN)_{\sigma(3)3} \ldots (MN)_{\sigma(n)n}$$

$$= \sum_{\sigma \in S_n} \varepsilon(\sigma) \sum_{k_1, k_2, k_3 \ldots k_n = 1}^{n} M_{\sigma(1)k_1} N_{k_1 1} M_{\sigma(2)k_2} N_{k_2 2} M_{\sigma(3)k_3} N_{k_3 3} \ldots M_{\sigma(n)k_n} N_{k_n n} =$$

$$\sum_{k_1, k_2, k_3 \ldots k_n = 1}^{n} \varepsilon(\sigma) N_{k_1 1} N_{k_2 2} N_{k_3 3} \ldots N_{k_n n} \sum_{\sigma \in S_n} M_{\sigma(1)k_1} M_{\sigma(2)k_2} M_{\sigma(3)k_3} \ldots M_{\sigma(n)k_n}$$

But the only terms that survive here are those where $k_1, k_2, k_3 \dots k_n$ are distinct, since otherwise we would be finding the determinant of a matrix with two equal columns. Now we have

$$\sum_{\rho \in S_n} \varepsilon(\sigma) N_{\rho(1)1} N_{\rho(2)2} N_{\rho(3)3} \dots N_{\rho(n)n} \sum_{\sigma \in S_n} M_{\sigma(1)\rho(1)} M_{\sigma(2)\rho(2)} M_{\sigma(3)\rho(3)} \dots M_{\sigma(n)\rho(n)}$$

$$= \det(N) \frac{\varepsilon(\sigma)}{\varepsilon(\rho)} \sum_{\sigma \in S_n} \varepsilon(\rho) M_{\sigma(1)1} M_{\sigma(2)2} M_{\sigma(3)3} \dots M_{\sigma(n)n} = \det(N)\det(M).$$

Now I will show that we can find the determinant by expanding by a row or column and taking sub-determinants.

Specifically, if Δ_{ij} is the determinant of M with i and j removed multiplied by $(-1)^{i+j}$, then $M_{ij}\Delta_{ij}$ with the summation convention on i and j fixed is equal to det(M).

Lemma: Suppose we have a matrix where if we remove row i and column j we get the matrix A, the ij entry of this matrix is 1, and all other entries in the i'th row and j'th column are 0, then its determinant is equal to $(-1)^{i+j}\det(A)$ and thus Δ_{ij}.

Proof: We can do i-1 row swaps and j-1 column swaps to turn this into a matrix with 1 in the top left and A after the bottom left corner of the 1. Then by a simple volume argument the determinant of this is equal to the determinant of A. (ie, in 3d, a parallelopiped if we keep the x axis fixed and move the y and z axis parallel to the x axis has the same volume scale factor equal to the area scale factor of the parallelogram scale factor of the y-z parallelogram). Then we have i+j-2 sign changes, and thus i+j sign changes, so this completes the proof of the lemma.

Theorem: $Det(M) = M_{ij}\Delta_{ij}$ with j fixed

Proof:

Notation: $[C_1, C_2, \dots C_n]$ will denote the determinant of the matrix with it's columns as $C_1, C_2, \dots C_n$.

$$Det(M) = [C_1, C_2, \dots, C_j, \dots, C_n]$$

Note that by matrix multiplication, $C_j = M_{ij}e_i$ with the summation convention. Therefore we can write

$Det(M) = [C_1, C_2, \dots, M_{ij}e_i, \dots, C_n] = M_{ij}[C_1, C_2, \dots, e_i, \dots, C_n]$ since the determinant is linear in each column from the level 6 video. $[C_1, C_2, \dots, e_i, \dots, C_n] = \Delta_{ij}$ since we can go and subtract e_i from each column until everything else in the i'th row is 0. So done.

Lecture 16:

Definition: The **adjugate** of a matrix M is $M^{-1}\det(M)$. If M is not invertible, then $Adj(M)$ is the transpose of the cofactor matrix, ie the thing you get when finding the inverse before dividing by the determinant. Note that we used this adjuagate matrix in the Cayley Hamilton Theorem proof.

Example: Consider the matrix $M := \begin{pmatrix} 1 & x & 1 \\ 1 & 1 & x \\ x & 1 & 1 \end{pmatrix}$. Lets try to compute Det(M).

Note that adding a multiple of 1 row to another does nothing to the determinant, so we can subtract the first row from the second and try to compute the determinant of $\begin{pmatrix} 1 & x & 1 \\ 0 & 1-x & x-1 \\ x & 1 & 1 \end{pmatrix}$. We can also subtract x times the first row from the third to get $\begin{pmatrix} 1 & x & 1 \\ 0 & 1-x & x-1 \\ 0 & 1-x^2 & 1-x \end{pmatrix}$. We have now reduced the problem to finding $\det\begin{pmatrix} 1-x & x-1 \\ 1-x^2 & 1-x \end{pmatrix}$ which is $x^3 - 3x + 2$. This is useful because if you try to "naively" compute the determinant of, say, a 5*5 matrix you could have up to 120 terms, and this just gets much worse for larger matrices.

We have seen from A level that we can use matrices to solve systems of linear equations.

Consider a system of n linear equations for unknowns $x_1, x_2, x_3, \dots x_n$. Now we rewrite the system in vector-matrix form as $Ax = b$. If A has non-zero determinant we know what to do and there will be a unique solution $x = A^{-1}b$. Therefore we will talk about what happens in the other case.

Case 1: b is not in the image of A in which case there are no solutions.

Case 2: b is in the image of A in which we have infinitely many solutions: In this case we take a particular solution and add any element in the kernel of A. If u is a solution to $Ax = b$, then x is a solution if and only if x-u is in the kernel of A, since we need $A(x-u) = 0$.

Example: Let x and y be real numbers, and we want to solve Ax+B with $A = \begin{pmatrix} 1 & x & 1 \\ 1 & 1 & x \\ x & 1 & 1 \end{pmatrix}, b = \begin{pmatrix} 1 \\ y \\ 1 \end{pmatrix}$,

$x \neq 1, x \neq -2$.

Now we can compute A^{-1}: $A^{-1} = \frac{1}{x^3-3x+2}\begin{pmatrix} 1-x & 1-x & x^2-1 \\ x^2-1 & 1-x & 1-x \\ 1-x & x^2-1 & 1-x \end{pmatrix}$. Now we can find by getting rid of certain common factors that our solution is

$$\frac{1}{x^2+x-2}\begin{pmatrix} -1 & -1 & x+1 \\ x+1 & -1 & -1 \\ -1 & x+1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ y \\ 1 \end{pmatrix} = \frac{1}{x^2+x-2}\begin{pmatrix} x-y \\ x-y \\ xy+y-2 \end{pmatrix}.$$

If x=1, then the kernel of A has dimension 2, so there is a plane of solutions if and only if y=1 (since y must be in the image of a which is multiples of (1,1,1)), otherwise there are none.

If x=-2, then the kernel of A has dimension 1, so there is a line of solutions if b is in the image of A. This turns out to be the case exactly when y=-2 – I won't go through this too carefully but essentially this is because $\begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ -2 & 1 & 1 \end{pmatrix}\begin{pmatrix} a-1 \\ a-1 \\ a \end{pmatrix}$ must be the form in order to get 1 in the first and third entry (we need the same first and second entries, and we need it to be such that they are both 1), and in any case we end up with -2 in the middle.

I notice that above the dimension of the kernel equals the multiplicity of the root in the determinant. I wonder if something more general along these lines can be said?

Lecture 17:

Later this lecture we will give a more efficient method of solving systems of linear equations since in practice inverting a matrix is very difficult and slow for anything larger than 3*3.

Consider the matrix equation $Au = 0$. This equation is saying $R_1.u = R_2.u = R_3.u = 0$, so we are looking at the intersection of 3 planes. But we knew this from A level, as well as the idea that if the right hand side is non-zero there may be no solutions, in which case the planes either form a sheaf or have two or all three parallel, and in higher dimensions I'm sure there are way more possible cases. The solution if the right hand side is 0 is by definition equal to the kernel of A, which is just 0 if and only if A is invertible.

Recall from A level (in deriving the eigenvalue stuff): Suppose there exists a non-zero vector v such that $Av = 0$, then since both $A0 = Av = 0$, this means that A is a many-to-one map so it is not invertible, so this implies that A has determinant 0. Conversely, if A is not invertible, then Av=0 has a non-zero solution, because A must have a kernel. We did not need this part for A level since we were only ever given matrices where we could actually find 3 linearly independent eigenvectors, and this is not always the case, but the fact that when this is the case there is an eigenvector is obvious and does not need proof. And now we know there is always an eigenvector. We will need to work in the complex numbers since a real matrix can have non-real eigenvalues from the characteristic equation, and at A level cases like this were avoided, but they can come up.

Note that if we go back to the intersection of the 3 planes idea, the dimension of the intersection of the three planes equals the dimension of the solution space which equals the dimension of the kernel. But then the number of linearly independent normals to these planes is 3 minus this, which equals the rank. This is hinting is a more general theorem which says that the dimension of the row space of any matrix equals the dimension of the column space. The lecturer did not do this but it is fundamental to linear algebra so I will.

Here is the proof from wikipedia, noting that orthogonal means perpendicular.

Let A be an $m \times n$ matrix with entries in the real numbers whose row rank is r. Therefore, the dimension of the row space of A is r. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r$ be a basis of the row space of A. We claim that the vectors $A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_r$ are linearly independent. To see why, consider a linear homogeneous relation involving these vectors with scalar coefficients $c_1, c_2, \ldots, c_r$:

$$0 = c_1 A\mathbf{x}_1 + c_2 A\mathbf{x}_2 + \cdots + c_r A\mathbf{x}_r = A(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_r\mathbf{x}_r) = A\mathbf{v},$$

where $\mathbf{v} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_r\mathbf{x}_r$. We make two observations: (a) $\mathbf{v}$ is a linear combination of vectors in the row space of A, which implies that $\mathbf{v}$ belongs to the row space of A, and (b) since $A\mathbf{v} = 0$, the vector $\mathbf{v}$ is orthogonal to every row vector of A and, hence, is orthogonal to every vector in the row space of A. The facts (a) and (b) together imply that $\mathbf{v}$ is orthogonal to itself, which proves that $\mathbf{v} = 0$ or, by the definition of $\mathbf{v}$,

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_r\mathbf{x}_r = 0.$$

But recall that the $\mathbf{x}_i$ were chosen as a basis of the row space of A and so are linearly independent. This implies that $c_1 = c_2 = \cdots = c_r = 0$. It follows that $A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_r$ are linearly independent.

Every $A\mathbf{x}_i$ is in the column space of A. So, $A\mathbf{x}_1, A\mathbf{x}_2, \ldots, A\mathbf{x}_r$ is a set of r linearly independent vectors in the column space of A and, hence, the dimension of the column space of A (i.e., the column rank of A) must be at least as big as r. This proves that row rank of A is no larger than the column rank of A. Now apply this result to the transpose of A to get the reverse inequality and conclude as in the previous proof.

Image: The proof from wikipedia that row rank = column rank for all matrices.

Now as promised I will show the easier method of finding the solutions to the system of equations. This method is called Gaussian elimination.

Consider a system of m equations with n unknowns, with m and n possibly different. Then we look for one coefficient that is non-zero (if this does not exist then this is not very interesting), and move it to the top left corner of the system.

Take the first equation of the form $A_1x_1 + A_2x_2 + A_3x_3 + \cdots A_kx_4 = b_j$ with the first term non-zero, then subtract multiples of it from the rest of the equations such that all their first coefficients are 0.

Now we have m-1 equations with at most n-1 non-vanishing unknowns, and we repeat the same procedure again. After repeatedly doing this, we have 1 equation with n unknowns, another one with n-1 unknowns, another one with n-2 and so on. If we have more unknowns than equations, then we will get incomplete information since every equation may have many unknowns, so there will be many solutions. If we have more equations than unknowns, or we get to a point that we cannot find another non-zero element, we will end up with a lot of 0=(linear combination of the b's) equations at the end. If these linear combinations are ever not 0, we have no solutions, but otherwise we can work backwards from the n'th equation to find all the unknowns.

Now we can interpret this another way: We are taking a matrix and by only interchanging rows and columns and adding multiples of rows to other rows we get it into row echelon form or upper triangular form, meaning every element below the main diagonal is zero. If m=n, then one can see that the determinant of the matrix in question is (possibly with a sign change) equal to the product of these diagonal entries.

In the coming lectures, we will explore more properties of Eigenvalues and Eigenvectors. We will only do this for real and complex matrices – for general linear maps between vector spaces that is beyond the scope of this course.

For example, now we know that the fundamental theorem of algebra holds, and this means that every matrix actually has Eigenvalues since the characteristic equation has (possibly complex) roots. At A level we knew this for real 3x3 and general 2x2 matrices – which was all we needed - but now we know it fully.

Definition: The **multiplicity** of a root k of a polynomial is the number of times you can factor out x-k. The multiplicity of an eigenvalue k is the multiplicity of k in the characteristic equation of the matrix.

Lecture 18:

Definition: A matrix is **diagonalizable** if it can be written as $M = PDP^{-1}$ for a diagonal matrix D.

Definition: The **algebraic multiplicity** of an eigenvalue λ is the multiplicity of λ as the root of the characteristic polynomial.

Definition: The **geometric multiplicity** of an eigenvalue λ is the dimension of the space of vectors v such that Mv=λv. This is called the eigenspace.

Proposition: By the fundamental theorem of algebra there are exactly n eigenvalues for an n*n matrix if we count multiple times for algebraic multiplicity.

Proposition: The determinant is the product of the eigenvalues with algebraic multiplicity as their power.

Intuition: If the matrix is really just taking a basis for the vector space and scaling it by eigenvalues, then the volume scale factor is clearly the product of those. But this only works for "nice" matrices.

Proof: Det(M)=Det(M-0I) which is the characteristic polynomial evaluated at 0. But this is just the constant term of the characteristic polynomial, which we know is the product of the roots (possibly with a minus sign – we will address this) which is the product of the eigenvalues.

About the sign change: The constant term of the characteristic polynomial is $(-1)^n\lambda^n$ so we can multiply the whole polynomial by $(-1)^n$ to get it to have 1 as the leading term, and then the sign change of the constant term is $(-1)^n$, but the constant term is also $(-1)^n$ times the product of the roots, so everything aligns correctly.

Sine polynomials with real coefficeints have roots that are real or come in conjugate pairs, this must hold for eigenvalues of real matrices.

Proposition: The trace of a matrix as defined as the sum of elements along the diagonal is actually the same as the sum of the eigenvalues.

Proof: Set the equation to $Det(\lambda I - M)$ so that 1 is the leading term In the polynomial. Then the sum of the roots is minus the second leading term. But we only get a second leading term from the determinant term given by $A_{11}A_{22} \dots A_{nn}$ where $A = \lambda I - M$. But $A_{11}A_{22} \dots A_{nn} = \prod_{i=1}^{n}(\lambda - M_{ii})$, which has minus the sum of the diagonal entries as its second leading terms. All other products in the determinant have to have no more than n-2 copies of λ so the λ^{n-1} term is fixed at this.

Example, since in A level we always had real eigenvalues, here is a case where we don't:

Consider $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then the characteristic equation is $\lambda^2 + 1 = 0$, so the eigenvalues are $\pm i$. Lets compute the eigenvectors associated with these. We want to solve (A-Ii)v=0: $\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Therefore we have that -ix-y=0 and the other equation is equivalent. So the eigenvectors for the eigenvalue i are all multiples of $\begin{pmatrix} 1 \\ -i \end{pmatrix}$. By the same calculation, for the eigenvalue -i the eigenvectors are all multiples of $\begin{pmatrix} 1 \\ i \end{pmatrix}$.

Example: The eigenvalues of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ are just 1 with algebraic multiplicity 2. But the geometric multiplicity is not 2 If we solve (A-I)v=0: $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ if and only if y=0, which means that the geometric multiplicity is 1 since the eigenvector is just the x axis.

Also, we cannot diagonalize $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, since the eigenvector matrix is just $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ which is I, but for any P, $PIP^{-1} = I$ and not $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ so this is a contradiction.

It turns out that a matrix is diagonalizable if and only if the geometric multiplicities are all equal to the algebraic multiplicities, and of course we will prove this.

Proposition: Eigenspaces are actually vector spaces and therefore geometric multiplicity is well defined.

Proof: The eigenspace is exactly the kernel of a matrix $M - \lambda I$. But kernels are subspaces since they are closed under additions and inverses and stuff (cf groups where we see that kernels are subgroups).

Proposition: The sum of geometric multiplicites is at most the dimension of the whole space because otherwise the dimension would be too high.

Proof: I will put a proof in the notes for next lecture due to some stuff to cover first (similar matrices). For now we will use it (ok since we promise to prove it in due course).

Definition: The **defecit** of an eigenvalue is the difference between their algebraic and geometric multiplicities. We will show also later in this course that the defecits are all 0 if and only if the matrix is diagonalizable.

Example: Consider $A = \begin{pmatrix} 4 & 1 & 0 \\ 0 & 4 & 1 \\ 0 & 0 & 4 \end{pmatrix}$. Then the characteristic polynomial is just $(4-\lambda)^3$.

Lets see what happens for the geometric multiplicity. We want (A-4I)v=0, so

$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0$, which implies that y=z=0, so the x axis is the eigenspace, so the geometric multiplicity is 1 so the defecit is 2.

Example: If we take a reflection in a plane with normal n, then geometrically we see that we have an eigenvalue 1 with eigenspace parallel to the plane and eigenvalue -1 with eigenspace parallel to the normal n.

Example: A rotation in 2 dimensions is given by $\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. The characteristic polynomial of this is $\lambda^2 - 2\lambda\cos(\theta) + 1 = 0$, and one checks that the eigenvalues are $e^{\pm i\theta}$ with eigenvectors parallel to $\begin{pmatrix} 1 \\ \mp i \end{pmatrix}$.

If we have a rotation in 3D about n, then we must have n an eigenvector with eigenvalue 1, and perpendicular to n we have the same behavior as in the 2D case, ie eigenvalues $e^{\pm i\theta}$ with eigenvectors $\begin{pmatrix} 1 \\ \mp i \end{pmatrix}$ where $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are considered to be perpendicular vectors that span the plane perpendicular to n.

Example: Consider $A = \begin{pmatrix} -3 & -1 & 1 \\ -1 & -3 & 1 \\ -2 & -2 & 0 \end{pmatrix}$. Then, I won't go through this explicitly, but it turns out that we have only an eigenvalue -2 with algebraic multiplicity 3. But we can see that the kernel of A+2I will not have dimension 3 since that would imply A+2I=0, so we do have a defecit. But A+2I is $\begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ -2 & -2 & 2 \end{pmatrix}$ which we see has rank 1 since it puts everything to a line. Therefore the geometric multiplicity has dimension of the nullity of that matrix which is 2.

Proposition: Diagonalizable (in the complex numbers) is equivelent to geometric and algebraic multiplicities being the same. In the real numbers this is the case if all eigenvalues are real.

Proof: We will do this next lecture due to time constraints.

Lecture 19:

Suppose the geometric and algebraic multiplicities are the same, then we have eigenvectors that form a basis for the vector space, and thus the matrix is diagonalizable.

If the matrix is diagonalizable, then it has to be that $A = PDP^{-1}$ and since P is invertible it must form a basis of eigenvectors so there has to be as much total geometric multiplicity as n.

Now we know that if eigenvalues are all distinct the matrix is diagonalizable. But the matrix can still be diagonalizable otherwise, it just isn't always.

The lecturer also provided the proof that a set of eigenvectors with distinct eigenvalues is linearly independent. See the proof of this in level 6.

Definition: Matrices A and B are **similar** if there exists a matrix P such that $B = P^{-1}AP$. Equivalently, they are similar if they are conjugate in the groups sense. We can now interpret a matrix as diagonalizable if it is similar to a diagonal matrix. In fact, similar matrices can be thought of as representing the same linear map with respect to a different basis. This is not too hard to see intuitively but we will look into this idea more formally soon.

Proposition: If matrices are similar then they have the same characteristic polynomial, and thus the same trace and determinant and eigenvalues.

Proof: $Det(B - \lambda I) = Det(P^{-1}AP - \lambda I) = Det(P^{-1}AP - P^{-1}(\lambda I)P) = Det(P^{-1}(A - \lambda I)P)$. Decomposing the determinant as a product we see that $Det(B - \lambda I) = Det(A - \lambda I)$. So done.

Theorem: Geometric multiplicity is at most algebraic multiplicity.

Proof: Let an eigenvalue have geometric multiplicity k in a matrix A. Then A is similar to a matrix that can be written as four blocks as follows:

$$\begin{pmatrix} \lambda I_k & B \\ 0 & C \end{pmatrix}$$

Now lets compute the characteristic polynomial of this, which must be the same as that of the matrix A:

$$P(x) = Det\begin{pmatrix} (\lambda - x) = I_k & B \\ 0 & C - xI_{n-k} \end{pmatrix}$$

Note that since we can do column operations (remove multiples of the first k columns until B vanishes) without changing the determinant, it is then easy to see (by the volume property) that $P(x) = (\lambda - x)^k Det(C - xI_{n-k})$. We have the factor $(\lambda - x)^k$ proving that the algebraic multiplicity is at least k, so done.

Theorem: The eigenvaleus of a hermitian matrix are real

Proof: Let v be an eigenvector with eigenvalue λ.

$v^\dagger(Av) = v^\dagger(A^\dagger v) = (v^\dagger A^\dagger)v = (Av)^\dagger v$ since A is hermitian so

$\lambda v^\dagger v = \bar{\lambda} v^\dagger v$

But λ equals its conjugate so is thus real.

Theorem: Eigenvectors of a hermitian matrix from distinct eigenvectors are orthogonal

Proof: Copy the level 6 proof about decomposition of symmetric matrices which proved the same thing, and just replace transpose with hermitian conjugate and everything accordingly.

Theorem: For each eigenvalue of a real symmetric matrix we can pick a real eigenvector v.

Proof: Let u be the real part of v. Then Au is real and A(v-u) is imaginary, and thus are the real and imaginary parts of Av, so Au=λu since the real parts match, so u is a real eigenvector.

Lecture 20:

Let $\{w_1, w_2, \dots, w_r\}$ be a set of r linearly independent vectors. We will construct a sequence of sets of the form $\{u_1, w_2', \dots, w_r'\}$, $\{u_1, u_2, w_3'', \dots, w_r''\}$ such that at the end we have a set of orthonormal vectors, such that each set has the same span and is linearly idnependent. To do this we will make sure the u_i's are orthonormal to eachother and orthogonal to the w's. In the first step we will simply take $u_1 := \frac{w_1}{|w_1|}$ and write $w_j' := w_j - (u_j . w_j)u_1$ to ensure u_1 has length 1 and is perpendicular to every other vector in the set. Now we do the same thing to turn $\{w_2', w_3', \dots, w_r'\}$ into $\{u_2, w_3'', \dots, w_r''\}$ until we finish, where at the end we will have an orthonormal basis for our vector space. We have the gurantee at each step that everything is orthogonal to previous u vectors by construction.

Definition: an **Eigenspace** is a space of eigenvectors corresponding to an eigenvalue.

We can now find an orthonormal basis for each eigenspace of a hermitian matrix A by the proposition above.

Theorem: Let A be a Hermitian matrix of size n*n, then A is diagonalizable, and in fact we can diagonalize it by $P^\dagger AP = D$ as that is equal to $P^{-1}AP$ since P can be chosen to be unitary by above. This is stronger than what was used at A level since this allows for repeated eigenvalues.

Proof: Induction on n: For n=1 this is trivial.

Now suppose it is true for n=k. Then any (k+1)x(k+1) hermitian matrix H has a characteristic polynomial which has a real root with an eigenvector. Now pick any eigenvector and pick a basis that includes this eigenvector. Now apply the above procedure (called Gram Schmidt) to turn it into an orthonormal basis. Now we will suppose that this is our basis – we will talk more on this point soon, but basically we know

$$He_1 = \lambda_1 e_1$$

Now we can write our matrix like

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \dots & \dots & \dots \\ \dots & \dots & C & \dots \\ 0 & \dots & \dots & \dots \end{pmatrix}$$

Proof that we have zeroes on the top row: The matrix Is hermitian, but it is not obvious since it is with respect to another basis. But don't worry – this basis is orthonormal. We can move it to the standard basis using a unitary matrix, do the transformation, then move it back. Ie, $A = UHU^{-1}$, where since U is unitary it equals its hermitian conjugate. $A = UHU^\dagger$, now use the fact that A is hermitian and take the conjugate of both sides to get that $A = A^\dagger$ follows. Now "C" is a k*k hermitian matrix with respect to an orthonormal basis so we can apply the induction hypothesis then we are done.

More on "change of basis" stuff: If A represents a linear map T from V with basis $\{e_1, \dots e_n\}$ to W with basis $\{f_1, \dots f_m\}$ then $T(e_i) = \sum f_j A_{ji}$. If bases of the same vector space are related by $e_i' = \sum e_k P_{ki}$ we can construct the matrix P accordingly to do a change of basis. If we also set $T(e_i') = \sum f_j' B_{ji}$ and $f_i' = \sum f_k Q_{ki}$, then (Proposition) we have a change of basis by $B = Q^{-1}AP$ where A and B represent the same linear map with respect to a different basis

Proof of the proposition:

$$T(e_i') = T(\sum e_k P_{ki}) = \sum T(e_k) P_{ki} = \sum \sum f_j A_{jk} P_{ki} = \sum f_j (AP)_{ji}$$

$$T(e_i') = \sum f_j' B_{ji} = \sum \sum f_k Q_{kj} B_{ji} = \sum f_k (QB)_{ki}$$

All steps directly from the definitions above, being careful to use the correct indices.

We now rename the indices so that they match and then we see that QB=AP by comparing coefficients so it follows that $B = Q^{-1}AP$. Q is invertible because everything is linearly independent.

Here is an intuitive diagram to show whatever the heck was just happening, noting that the writing order of matrix multiplication is the reverse of the order the maps happen in.

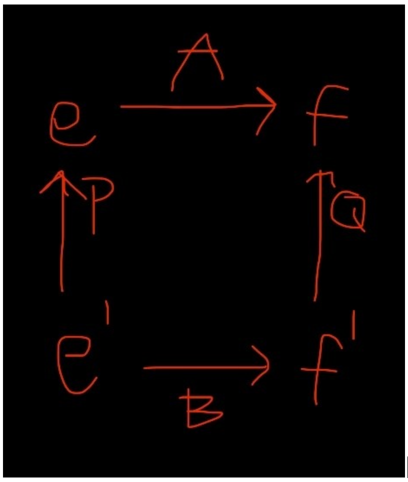


Image: The promised diagram

When we left multiply a vector by P (say Pa=b), we note that the columns of P are how $e_i{}'$ is formed by a linear combination of e_j, which means if our vector a is actually how we write the vector in e' coordinates (basis) then the result b is how we write it in the original e coordinates (again it's a linear combination of components), as b_k comes from the contribution of how much the a components add to the k'th component of b, which is based on the k'th column of P. This means (we can say) P sends e' to e (compare with matrix A, representing the linear transformation. we write some vector in coordinate e, multiply by a, and get the result written in coordinates f, meaning we send e to f)

Remark: Column i of A represents $T(e_i)$ with respect to the f basis. This is a generalization of a known fact about matrices from A level.

If we want to change basis in the other direction, then the matrices P', Q' we need are such that $P' = P^{-1}$ and $Q' = Q^{-1}$, since we want $A = Q'^{-1}BP'$.

Example: Consider dim(V)=2, dim(W)=3 and

$$T(e_1) = f_1 + 2f_2 - f_3$$

$$T(e_2) = -f_1 + 2f_2 + f_3$$

Then we can write the matrix A as

$$\begin{pmatrix} 1 & -1 \\ 2 & 2 \\ -1 & 1 \end{pmatrix}$$

Now consider a basis such that $e_1' = e_1 - e_2, e_2' = e_1 + e_2$

This makes $P = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$.

Now let w be such that $f_1' = f_1 - f_3, f_2' = f_2, f_3' = f_1 + f_3$, this makes

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

Therefore the change of basis formula gives

$B = Q^{-1}AP = \begin{pmatrix} 2 & 0 \\ 0 & 4 \\ 0 & 0 \end{pmatrix}$. Therefore $T(e_1') = 2f_1{}'$, $T(e_2') = 4f_2{}'$, $T(e_3') = 0$.

Note: If V=W with the same basis then we must have P=Q. Therefore, matrices represent the same linear map $T: V \to V$ if and only if they are similar.

Lecture 21:

Here we reprove the change of basis formula just to make sure it's really clear, I guess (I'm not sure why we're re-proving it):

Consider a vector space V and x a vector in V, and let V have 2 bases $\{e_i\}, \{e_i'\}$ related by a matrix P by $e_j' = \sum e_i P_{ij}$ and $x = \sum x_i e_i = \sum x_i' e_i'$. But now we know that $x = \sum_j x_j' e_j' = \sum_j x_j' \sum_i e_i P_{ij}$. Therefore

$\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} P_{11} & \dots & \dots & P_{1n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ P_{n1} & \dots & \dots & P_{nn} \end{pmatrix} \begin{pmatrix} x_1{}' \\ x_2{}' \\ \dots \\ x_n{}' \end{pmatrix}$, therefore x=Px'.

Similarly, consider a vector space W with a vector y and bases $\{f_i\}, \{f_i'\}$, then by the same argument y=Qy'. If we have a linear map V to W then we can write it in matrix form as y=Ax and y'=Bx'. Combining this with what we proved above we get the same conclusion: $B = Q^{-1}AP$.

On the Cayley-Hamilton theorem:

We can verify it for matrices of a specific size (like 2*2 or 3*3) by just doing some tedious algebra. We won't do this but you can do it yourself if you want to, but of course this is not possible to do for infinitely many sizes.

If A is diagonalizable, the proof is much easier. $P^{-1}AP = D$ and the characteristic polynomial applied to D gives 0 since its diagonal entries are roots of this equation. To spell out why this extends to A,

$$P^{-1}APP^{-1}APP^{-1}AP \dots P^{-1}AP = P^{-1}AAAA \dots AP = P^{-1}A^kP = D^k$$

$$A = P0P^{-1} = 0$$

We can argue that any matrix is arbitrarily close to a diagonalizable matrix if we perturb it in a clever way to get a second proof different from the one in Level 6 (which the lecturer is doing but this is non-examinable), but the details can be tricky and this is beyond the course.

Definition: A quadratic form is a function $F: \mathbb{R}^n \to \mathbb{R}$ such that $F(x) = x^TAx$ where A is a real symmetric matrix. We can diagonalize A to get $F(x) = (P^Tx)^TD(P^Tx)$. We will rename P^Tx to x'. If we set $x = x_1e_1 + \cdots x_ne_n, x' = x_1'e_1 + \cdots + x_n'e_n$ then we can make a new basis such that

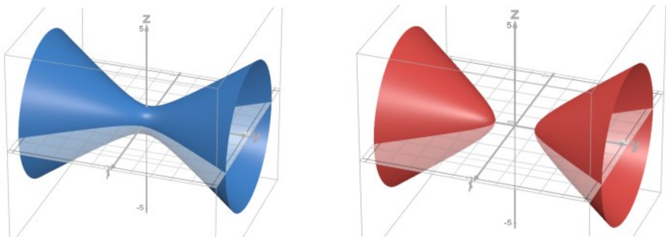
$x = x_1{}'u_1 + \cdots x_n{}'u_n$, and we will call this the principal axis of the quadratic form. These are related to the standard axis by the orthogonal matrix P. Because of this, $|x|^2 = \sum x_i x_i = \sum x_i' x_i'$.

Example, let $F(x) = x^TAx$ with $A = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$. This has eigenvalues $a \pm b$ and normalized eigenvectors $\frac{1}{\sqrt{2}}\begin{pmatrix} \pm 1 \\ 1 \end{pmatrix}$. We can see that $F(x) = ax_1^2 + 2bx_1x_2 + ax_2^2 = (a+b)(x_1')^2 + (a-b)(x_2')^2$.

As an example, if we take $a = \frac{3}{2}, b = -\frac{1}{2}$ then $F(x) = (x_1')^2 + 2(x_2')^2$ which is an ellipse with axis being at the eigenvectors. If we take $a = -\frac{1}{2}, b = \frac{3}{2}$ then we get $F(x) = (x_1')^2 - 2(x_2')^2$ which is a hyperbola.

Lecture 22:

Example: Let F(x) be a quadratic form that is already diagonalized so we can write $F(x) = x^T D x$ with respect to our principal axis. If the eigenvalues are all positive then we set F to equal a constant we get an ellipsoid, ie a sphere with axes stretched. If eigenvalues are some positive and some negative, like in the case $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, we get a hyperboloid, which could be either in one piece or two pieces, I will show an image to visualise this:



Images: 1-sheated vs 2-sheated hyperbola.

Note that a matrix M can be decomposed additively into a symmetric and antisymmetric part. For A antisymmetric, one checks that $x^T A x = 0$, and this is why we only define this for symmetric matrices.

Defintion: A **quadric** in $\mathbb{R}^n$ is a hypersurface defined by setting $Q(x) \coloneqq x^T A x + b^T x + c = 0$ for a symmetric n*n real matrix A and b a vector in $\mathbb{R}^n$.

We want to classify this up to solutions related by rotations and reflections and translations.

If A is invertible then we can complete the square. We can take a vector $y = x + \frac{1}{2}A^{-1}b$. If we do algebra on this we see that $y^T = x^T + \frac{1}{2}b^T A^{-1}$, and that $y^T A y = x^T A x + \frac{1}{2}b^T x + \frac{1}{4}b^T A^{-1} b$. What we can do is put $y^T A y = x^T A x + \frac{1}{2}b^T x + \frac{1}{4}b^T A^{-1} b + c - c = Q(x) + \frac{1}{4}b^T A^{-1} b - c$. This means that $F(y) = \frac{1}{4}b^T A^{-1} b - c$ is equivalent to $Q(x) = 0$. Now we diagonalise F, then the eigenvalues of A and the value of $\frac{1}{4}b^T A^{-1} b - c$ are what determine the geometrical nature of the quadric. If they are all positive we get an ellipsoid, if the eigenvalues have different signs and $\frac{1}{4}b^T A^{-1} b - c \neq 0$ will produce a hyperboloid. If some eigenvalues are 0, then we will not be able to do the trick above and get linear and quadratic terms.

Definition: A **conic** is a quadric in $\mathbb{R}^2$. If A is invertible then we get $ax^2 + by^2 = c$ where x and y are renamed to be the principal axis, and this is an ellipse or a hyperbola or a point (if c=0) or nothing (if a,b>0 and c<0 or the other way around). If A is not invertible then we can still diagonalize it since it is symmetric, then we get $\lambda_1 x^2 + b_1 x + b_2 y + c = 0$ where x and y are our principal axes. This is a parabola (unless $b_2 = 0$ in which case it is a pair of lines or a line or nothing, or both eigenvalues are 0 in which case it is a single line). See level 6 for the half-visual-half-algebraic proof that all of these (except for the degenerate cases of nothing, points or lines) are actually the slices of a cone. I now am going to make a guess that quadrics in general are slices of higher dimensional cones but I don't know if this is true.

Lecture 23:

Now we redo everything about conic sections in A level further maths and in the level 6 thing on conic section properties. So look at that. I'm not writing up this conic eccentricity focus directrix parabola hyperbola ellipse nonsense again because I find it so boring. Moving on.

Theorem: Consider a matrix A of size 2*2 corresponding to a linear map from $\mathbb{C}^2 \to \mathbb{C}^2$. Then it is similar to one of the following:

i) $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ where there are no constraints on a and b – they may be 0 ot not or the same or not.

ii) $\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$

Proof: The characteristic polynomial of A has 2 roots counting multiplicity over $\mathbb{C}$. If the roots are distinct then A is diagonalizable so there is nothing to prove. In fact the only case we need to worry about is when we have a repeated eigenvalue with geometric multiplicity 1. Let v be an eigenvector with the eigenvalue which is λ and extend it to a basis by another linearly independent vector w. We know that $Av = \lambda v, Aw = av + bw$ since v and w are a basis. Therefore the matrix is similar to a matrix of the form $\begin{pmatrix} \lambda & a \\ 0 & b \end{pmatrix}$. Note that $b = \lambda$ since the characteristic polynomial must agree. So we are similar to a matrix of the form $\begin{pmatrix} \lambda & a \\ 0 & \lambda \end{pmatrix}$. Also, $a \neq 0$ since we do not have a diagonal matrix by assumption. We will now define $u = av$, then $Aw = u + \lambda w$. Now with respect to the basis {u, w} our matrix is $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. So done.

We now want to find a nice form for any matrix similar to any non-diagonalizable matrix to generalize the theorem above. This will take a lot of work but we will get a very useful result.

Definition: Eigenspace

The eigenspace of an eigenvector λ is exactly what you think it is – the vector space formed by eigenvectors, or the kernel of $A - \lambda I$. We write this as $E(\lambda)$

Definition (Direct sum): An internal direct sum of vector spaces $E_1 \oplus E_2 \oplus E_3 \dots E_{n-1} \oplus E_n$ is the vector space defined as a linear combination of elements of the E's where the E's are a subset of a vector space. This is only defined if any set of vectors from each of the spaces are linearly independent. An external direct sum is when we build a vector space from existing vector spaces instead of subspaces of an existing vector space. We will just need the internal one for this thing we're doing to generalize the theorem above.

We know that eigenspaces form an internal direct sum as we proved earlier that eigenvectors with distinct eigenvalues are linearly independent.

Proposition: A matrix A is diagonalizable in $\mathbb{C}$ if and only if there exists a non-zero polynomial p such that p(A)=0 and p(x) has no repeated roots.

Proof: If A is diagonalizable then the vector space is the direct sum of the eigenspaces because of geometric multiplicities and stuff. So v can be written uniquely as a linear combination of eigenvectors. Now consider the polyno mial $p(t) = \prod_{i=1}^{k}(t - \lambda_i)$ where k is the number of distinct eigenvalues of A. Then $p(A)v = \sum p(A)v_i = \sum p(\lambda_i)v_i = 0$ as again, A is effetively $\lambda_i I$ when it is acting on the vector v_i. Conversely, suppose such a polynomial exists. Then there must be one of the form $\prod_{i=1}^{k}(t - \lambda_i)$ with eigenvalues, since if we had another polynomial with p(A)=0 we could factor out non-eigenvalue factors by taking the inverse and multiplying. So suppose a polynomial of that form exists such that p(A)=0. Then we want to show that any vector in V is a sum of eigenvectors of distinct eigenvalues. Let $q_j(t) = \prod_{i \neq j} \frac{t-\lambda_i}{\lambda_j - \lambda_i}$ which is a polynomial of degree k-1. Note that $q_j(\lambda_i) = \delta_{ij}$. Now consider $q(t) = \sum_{i=1}^{k} q_i(t)$. Then $deg(q-1) < k$ but $q(\lambda_i) - 1 = 0$ for all I from 1 to k. Therefore q=1. Now let π_j be a matrix given by $q_j(A)$, then the above says that $\sum \pi_j = I$. Therefore given v in our vector space, we know that $v = \sum \pi_j\, v$. But then we want to show that $\pi_j v$ is in $E(\lambda_j)$, and this is true as we can reverse the definitions to get that $(A - \lambda_j I)\pi_j v = \frac{1}{\prod_{i \neq j} \lambda_j - \lambda_i} \left(\prod_{i=1}^{k} A - \lambda_i\right) v = \frac{1}{\prod_{i \neq j} \lambda_j - \lambda_i} p(a) v = 0$. Therefore $v = \sum \pi_j\, v$ is a sum of eigenvalues. So done. Note that in the proof above, π_j can be thought of as a projection onto the j'th Eigenspace.

Definition: The **minimal polynomial** of a matrix A is the polynomial p of least degree such that p(A)=0. This always exists by the cayley hamilton theorem. Note that if there are two minimal polynomials of the same degree that are not a constant multiple of eachother then we can rescale and subtract them in such a way that we get a polynomial of a smaller degree. So it is unique. Every polynomial p with p(A)=0 is a multiple of the minimal polynomial because we can write it as a multiple of the minimal polynomial plus a remainder R where p(R) cannot be 0.

Example: The minimal polynomial of I is t-1. The minimal polynomial of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is $(t-1)^2$: It is not linear but it divides the characteristic polynomial so there's not much else it can be since its degree is 2.

Theorem: A matrix is diagonalizable if and only if its minimal polynomial has no repeated roots

Proof: This follows from the previous theorem: If there is any polynomial with no repeated roots that the matrix "satisfies" then the minimal polynomial has no repeated roots, and the converse is true, so this statement is equivalent to the matrix being diagonalizable.

The multiplicity of an eigenvalue in the minimal polynomial which we write as c_λ gives a third multiplicity type. We will call the algebraic and geometric multiplicities a_λ and g_λ.

Lemma: Similar to for geometric multiplicity, $1 \leq c_\lambda \leq a_\lambda$

Proof: The second inequality is easy as the minimal polynomial divides the characteristic polynomial. The first inequality is because the minimal polynomial of a matrix applied to an eigenvector must send the eigenvector to 0, but because of the idea that a matrix can be considered to be λI when we are thinking about what it does to an eigenvector, it means p(λI)=0 so λ is a root of p.

Definition: We say a matrix is in **jordan normal form** if it is a block diagonal matrix of the form in the image below

$$\begin{pmatrix} J_{n_1}(\lambda_1) & & & 0 \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ 0 & & & J_{n_k}(\lambda_k) \end{pmatrix}$$

Where each "jordan block" is something like the image below, and the n's give the size of the corresponding matrix to the image below.

$$\begin{pmatrix} \lambda & 1 & \cdots & 0 \\ 0 & \lambda & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & \lambda \end{pmatrix}$$

So a matrix in jordan normal form might look something like the image below

$$\begin{pmatrix} \lambda_1 & 1 & & & & & & & \\ & \lambda_1 & 1 & & & & & & \\ & & \lambda_1 & 0 & & & & & \\ & & & \lambda_2 & 0 & & & & \\ & & & & \lambda_3 & 1 & & & \\ & & & & & \lambda_3 & 0 & & \\ & & & & & & \ddots & \ddots & \\ & & & & & & & \lambda_n & 1 \\ & & & & & & & & \lambda_n \end{pmatrix}$$

Theorem: Every matrix can be written in Jordan Normal Form in a unique way up to permutation of the blocks. We will talk about matrices that have jordan normal forms and eventually prove that every matrix does. The image below shows the possible jordan normal forms for 3*3 matrices, as well as their minimal and characteristic polynomials which we can work out (although for higher dimesnional matrices, even knowing both of these polynomials does not determine the normal form).

Jordan normal form	χ_A	M_A
$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$	$(t-\lambda_1)(t-\lambda_2)(t-\lambda_3)$	$(t-\lambda_1)(t-\lambda_2)(t-\lambda_3)$
$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}$	$(t-\lambda_1)^2(t-\lambda_2)$	$(t-\lambda_1)(t-\lambda_2)$
$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}$	$(t-\lambda_1)^2(t-\lambda_2)$	$(t-\lambda_1)^2(t-\lambda_2)$
$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix}$	$(t-\lambda_1)^3$	$(t-\lambda_1)$
$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix}$	$(t-\lambda_1)^3$	$(t-\lambda_1)^2$
$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix}$	$(t-\lambda_1)^3$	$(t-\lambda_1)^3$

Now note that $J_n(\lambda) = J_n(0) + I\lambda$

Note that in the standard basis, we have that $J_n(0)(e_1) = 0, J_n(0)(e_{i+1}) = e_i$. Therefore we know by considering what doing this several times would do that $(J_n(\lambda) - I\lambda)^k = \left(J_n(0)\right)^k = \begin{pmatrix} 0 & I_{n-k} \\ 0 & 0 \end{pmatrix}$. And if k>n then we get the 0 matrix. Therefore the minimal polynomial of $J_n(0)$ is t^n (this is minimal as any factor of this will not give 0), and thus the minimal polynomial of $J_n(\lambda)$ is $(t-\lambda)^n$ as if there was one with smaller degree we could shift it to get a smaller polynomial for $J_n(0)$. Let n(A) denote the dimension of the kernel of A, then $n(A) = \sum n(J_{n_i})$. It should hopefully be clear that

$n((J_m(\lambda) - \lambda I_m)^r) = \min(r, m)$. The intuition for this first m times we multiply this we essentially kill another column and increase the nullity by 1, then when the nullity is m we can't go any further.

Note that since blocks in block matrices all act on independent vector spaces, there are several things we can say. The determinant of the matrix is the product of the determinants of the blocks by a volume argument if we think about the standard basis. Therefore the characteristic polynomial is the product of the characteristic polynomial of the blocks. By definition of the minimal polynomial, the minimal polynomial is the lowest common multiple of the minimal polynomials of the blocks. Therefore because of these facts we know that g_λ is the number of jordan blocks with eigenvalue λ (because each jordan block of λ adds 1 to the size of the kernel of A-λI), a_λ is the sum of sizes of jordan blocks with eigenvalue λ, and c_λ is the size of the largest such block.

Lemma: Jordan normal forms are unique up to permuting the blocks if they exist

Proof: Suppose A is a matrix in jordan normal form. Then the number of jordan blocks for an eigenvalue λ that have at least size r is given by $n((A - \lambda I)^r) - n((a - \lambda I)^{r-1})$: The reason why is because it is exactly for the blocks of size at least r that multiplying for the r'th time increases the nullity by 1. But now we can work out by doing the right subtraction the number of jordan blocks of a certain size for a certain eigenvalue so they are indeed unique.

Back to the existence proof.

Definition: A **generalized eigenspace** is $V_i \coloneqq Ker((a - \lambda_i I)^{c_{\lambda_i}})$

Lemma: V is the direct sum of generalized eigenspaces, ie linear combinations of them span V and are linearly independent.

Proof: $p_j(t) \coloneqq \prod_{i \neq j}(t - \lambda_i)^{c_{\lambda_i}}$. Note that the highest common factor of p_1, p_2 is $\prod_{i=3}^{k}(t - \lambda_i)^{c_{\lambda_i}}$, so Bezout's identity for polynomials (Level 4) means $\prod_{i=3}^{k}(t - \lambda_i)^{c_{\lambda_i}}$ is a linear combination of p_1, p_2. But then $\prod_{i=4}^{k}(t - \lambda_i)^{c_{\lambda_i}}$ is the highest common factor of $\prod_{i=3}^{k}(t - \lambda_i)^{c_{\lambda_i}}$ and p_3 and is thus a linear combination of p_1, p_2, p_3. By repeated similar logic, or induction, whatever you want to call it, we have q polynomials such that $\sum p_i q_i = 1$. This is actually always true for coprime sets of polynomials and numbers by similar logic so this is a useful idea to keep in mind for the future and not just this proof. We now define the map $\pi_j \coloneqq p_j(A)q_j(A) = q_j(A)p_j(A)$ (since factorizations like this of a matrix commute as we can expand them the same regardless of the order). Then by construction, $\sum \pi_j = I$. We will now call the minimal polynomial of A M. Then M(A)=0 by definition but we know $M(t) = (t - \lambda_j I)^{c_{\lambda_j}} p_j(t)$, therefore $0 = M(A)q_j(A) = (t - \lambda_j I)^{c_{\lambda_j}} \pi_j(A)$, so the image of π_j is in the generalized eigenspace V_j. Now suppose v is in our vector space V, then $v = Iv = \sum \pi_j\, v$ which is in the set of linear combinations of stuff in the generalized eigenspaces. To show that this is a direct sum, note that $\pi_i \pi_j = 0$ since the product contains M(A) as a factor. $\pi_i = I\pi_i = (\sum \pi_j)\pi_i = \pi_i^2$, so π_i is a projection onto some space, and this space we know is contained in V_i, and it is V_i because $\pi_i v_i = v_i$. The reason is because

i) If v is in a different eigenspace, such as V_j, then by definition $(a - \lambda_j I)^{c_{\lambda_j}} v = 0$, but then π_i contains $(a - \lambda_j I)^{c_{\lambda_j}}$ as a factor so if we move that factor to the right we have that $\pi_i v = 0$

ii) If v is in V_i then $v = Iv = \left(\sum p_j q_j\right)v = \sum \pi_j v = \pi_i v$ since the other terms vanish. So done.

Therefore by the projection property if a non-trivial linear combination of vectors in the V_i's is 0, then doing π_i will make that 0 for all i, which will extract each component and show that each component is 0. So V is the direct sum as claimed.

We will now check that anything in a generalized eigenspace stays there under A. The proof is straight forward if we use the idea since we know that we can commute linear factors in a polynomial of the matrix: $(A - \lambda_i I)^{c_{\lambda_i}}(Av) = A(A - \lambda_i I)^{c_{\lambda_i}}(v) = A0 = 0$, so Av is in the generalized eigenspace by definition, so if we make the basis around the generalized eigenspaces then the matrix must be block diagonal because of this. Now we will check that any of these blocks in this block diagonal matrix only has 1 eigenvalue. The reason is because if $(A - \mu)v = 0$ where we are only working in this specific generalized eigenspace. But then by definition of the eigenspace, $(A - \lambda_i I)^{c_{\lambda_i}} v = 0$. Then $(A - \lambda_i I)v = (\mu - \lambda_i)v$ if v is an eigenvector with eigenvalue μ, so $0 = (A - \lambda_i I)^{c_{\lambda_i}} v = (\mu - \lambda_i)^{c_{\lambda_i}} v$, but v is not 0 so the only way this can happen is if $\mu = \lambda_i$ since this is just a constant that we are multiplying v by, so there is only one eigenvalue. Therefore if we can show the main theorem for matrices with only one eigenvalue, we are done. In fact, by subtracting λI, we can just show it for matrices where all eigenvalues are 0. Such matrices are called **Nilpotent**. Nilpotent is usually defined a different way but we will show that that definition is equivalent.

Proposition: A matrix is Nilpotent if some power of it is the zero matrix. This is equivalent to the above definition.

Proof: If all eigenvalues are 0 the other defintion follows from the cayley hamilton theorem. Conversely, suppose the matrix satisfies this other definition, then if an eigenvector has a non-zero eigenvalue then it will never go to 0 no matter how many times we multiply by the matrix, so all eigenvalues are 0. So done.

So now we just need one more thing.

Lemma: A nilpotent matrix is similar to a matrix in Jordan normal form.

Proof: Now suppose we have a nilpotent n*n matrix L with minimal polynomial t^k with $k > 1$ (since if k=1 the theorem is trivial). We see that the eigenvalue 0 has algebraic multiplicity n and geometric multiplicity n(L). L is not diagonalizable because otherwise it would be similar to the zero diagonal matrix and thus would equal zero. The images of L^n as n increases from 0 to k form a subset chain, ie $V \supseteq Im(L) \supseteq Im(L^2) \ldots \supseteq Im(L^k) = 0$. This is because $Im(L^{a+1}) \subseteq Im(L^a \circ L) \subseteq Im(L^a)$. These inclusions are actually strict since otherwise L would be a bijection on a non-zero vector space which is not possible since L is nilpotent. Therefore we have (by the rank nullity theorem, and the fact that a kernel of something is in a kernel of L times that thing) $0 \subset Ker(L) \subset Ker(L^2) \ldots \subset Ker(L^k) = V$. Let $\Omega_j = Ker(L) \cap Im(L^j)$, then it is easy to see that $Ker(L) = \Omega_0 \supseteq \Omega_1 \supseteq \Omega_2 \ldots \supseteq \Omega_k = 0$. Now lets think about the rank of the product of two matrices: Rank(XY) is gonna have to be the dimension of the image of XY, which is the dimension of the image of Y minus the null space of X when restricted to the image of Y, which is Ker(X)∩Im(Y). Therefore we get the formula Rank(XY)=Rank(Y)-Dim(Ker(X)∩Im(Y)). Therefore $d_j \coloneqq Dim(\Omega_j) = Rank(L^j) - Rank(L^{j+1})$ with $d_k = 0$. Now lets investigate Ω_{k-1}: Anything in the image of L^{k-1} will be sent to 0 if it is multiplied by L again, therefore it is in the kernel of L, so the intersection definition reduces to $\Omega_{k-1} = Im(L^{k-1})$. We can obtain a basis for Ω_{k-1}. We will write this basis as $w_1, w_2, \ldots w_{s_{k-1}}$ such that for each w_i I can find a vector $x_i^{(k-1)}$ such that $L^{k-1} x_i^{(k-1)} = w_i$. Now recall that $\Omega_{k-2} = Ker(L) \cap Im(L^{k-2})$. We can now extend the basis for Ω_{k-1} to a basis for Ω_{k-2}. Lets write this new basis as $w'_1, w'_2, \dots w'_{s_{k-2}}$. Since each thing is in $Im(L^{k-2})$, I can find x's such that we have that $L^{k-2}x_i^{(k-2)} = w'_i$. Now note that we must have that $s_j = d_j - d_{j+1}$ from how we have been using the s's. We will continue doing the process we have been doing to obtain a basis for $Ker(L)$.

Now we want to construct jordan chains: We had that, eg, $L^{k-2}x_i^{(k-2)} = w'_i$. The jordan chain is the following list of lists of vectors, and there is a jordan chain for each power of L. It's a lot:

$$L^{k-2}x_i^{(k-2)} = w'_i$$

$$L^{k-3}x_i^{(k-2)}$$

$$\dots$$

$$Lx_i^{(k-2)}$$

$$x_i^{(k-2)}$$

For i (in this example) ranging from 1 to s_{k-2}. I now claim that, in fact, the set of all k jordan chains forms a basis for V. From the way I have written this out it seems like there are a ton of these vectors compared to what you would expect the dimension of V to be, but it works as in practice the s's are often 0 so these don't all come up. Therefore we need to check that there are n of these vectors and that they are linearly independent. The total is $ks_{k-1} + (k-1)s_{k-2} + (k-2)s_{k-3} + \dots + 2s_1 + s_0$ since we are adding up the size of the jordan chains. Lets manipulate this sum a bit, keeping in mind that $d_k = 0$:

$$\sum_{j=0}^{k-1}(j+1)s_j = \sum_{j=0}^{k-1}(j+1)(d_j - d_{j+1}) = \sum_{j=0}^{k-1}(j+1)(d_j) - \sum_{j=0}^{k-1}(j+1)(d_{j+1})$$
$$= d_0 + \sum_{j=1}^{k-1}(j+1)(d_j) - \sum_{j=1}^{k-1}(j)(d_j) = d_0 + \sum_{j=1}^{k-1}(d_j)$$
$$= N(L) + \sum_{j=1}^{k-1} Rank(L^j) - Rank(L^{j+1}) = N(L) + Rank(L) - Rank(L^k) = n$$

By the method of differences, the rank-nullity theorem, and the fact that $L^k = 0$. So to check that we have a basis for V we just need to check that our vectors are linearly independent. We will now construct a few matrices:

Q_0 with the columns:

$$x_1^{(k-1)}, x_2^{(k-1)}, \dots, x_{s_{k-1}}^{(k-1)}$$

Q_1 with the columns:

$$Lx_1^{(k-1)}, Lx_2^{(k-1)}, \dots, Lx_{s_{k-1}}^{(k-1)}, x_1^{(k-2)}, x_2^{(k-2)}, \dots, x_{s_{k-2}}^{(k-2)}$$

Q_2 with the columns:

$$L^2x_1^{(k-1)}, L^2x_2^{(k-1)}, \dots, L^2x_{s_{k-1}}^{(k-1)}, Lx_1^{(k-2)}, Lx_2^{(k-2)}, \dots, Lx_{s_{k-2}}^{(k-2)}, x_1^{(k-3)}, x_2^{(k-3)}, \dots, x_{s_{k-3}}^{(k-3)}$$

And so on until Q_{k-1}.

Now we define a matrix Q to have columns $Q_{k-1}, Q_{k-2}, \dots, Q_2, Q_1, Q_0$. Now we will ask ourselves what the solution is to Qz=0 for vectors z: We hope that this is just z=0 since this would verify linear independence.

We will write $z = \begin{bmatrix} \dots \\ \dots \\ \bar{\bar{z}} \\ \bar{z} \\ \tilde{z} \end{bmatrix}$ where each thingy here is a sub-vector corresponding to each Q_i. We will now multiply the equation we are looking at on the left to get that $L^{k-1}Qz = 0$. By nilpotence, any column in Q that involves an L will go to 0 under the matrix $L^{k-1}Q$. But then consider what happens with Q_0: The vectors $x_1^{(k-2)}, x_2^{(k-2)}, \dots, x_{s_{k-1}}^{(k-1)}$ when multiplied by L^{k-1} are in the Ω's, in fact one of the w's from earlier and thus in the kernel of L and thus when we multiply them by we get 0. But with Q_1 and higher, we get a power of L above where we would be in the kernel and thus go to 0. Because the w's were part of a basis and were linearly independent, it means that if $L^{k-1}Q_0 z = 0$ implies z=0. Therefore, we know now that $\tilde{z} = 0$. Now we will look at $L^{k-2}Q_1 z = 0$, then by the same argument the only non vanishing parts are Q_0 and Q_1, but we are interested in $L^{k-2}Q_1$ as those columns are the (w')'s, which are again linearly independent by construction, so by the same argument we know that $\bar{z} = 0$. We can continue to do this and then we will get that z=0 so we indeed have a basis for V. We will now consider j ranging from 0 to k-1 v ranging from 1 to s_j and the j+1 column matrix of columns $\left[L^j x_v^{(j)}, L^{j-1} x_v^{(j)}, \dots, L x_v^{(j)}, x_v^{(j)}\right] := P_{j,v}$. Then $LP_{j,v} = \left[0, L^j x_v^{(j)}, \dots, L^2 x_v^{(j)}, L x_v^{(j)}\right]$ since$L^j x_v^{(j)}$ is in the kernel of L. What happens now is $P_{j,v} \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} = LP_{j,v}$ so we finally see some jordan stuff happening. We will make P out of all possible $P_{j,v}$'s, then we will achieve our goal because $P^{-1}LP$ is in jordan normal form: it is made out of the blocks corresponding to those 0-1 matrices we saw above. So done.

Proposition: All of the following are equivalent to a matrix R being orthogonal:

i) $(Rx).(Ry) = (x.y)$
ii) $R^T R = RR^T = I$
iii) $|Rx| = x$ for all vectors x
iv) Columns of R are orthonormal

Proof: See Groups notes – we just did this proof in that course. We define the special orthogonal group here but we just defined it in the groups course as well so please see that.

We can think of O(n) as preserving lengths and angles and SO(n) as also preserving orientations.

Lecture 24:

For a rotation matrix R we can make a new orthonormal basis by $u_i = \sum_j R_{ij} e_j$. We can think of it either as a change of basis or a transformation of vectors. The components of Rx with respect to the changed basis are the same as the components of x with respect to the normal basis. We can also think that a vector stays the same but the axes move in the opposite direction, ie by R^{-1}. Consider an "inner product" given by $(x, y) = x^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} y$. This not satisfy the properties that an inner product should satisfy because it is not positive definite, ie (x,x) is not necessarily a non-negative real number. However, the oother properties are satisfied if we are working in the real numbers. We still have that $(1,0) \coloneqq e_0$ and $(0,1) \coloneqq e_1$ are orthonormal in the sense that their product is 0, and their product with themselves are 1 and -1 respectively. This "inner product" is called the **Minkowski metric** and $\mathbb{R}^2$ equipped with this metric is called a **Minkowski space**.

Note that a linear map $T: \mathbb{R}^2 \to \mathbb{R}^2$ preserves the Minkowski metric if and only if $(Tx, Ty) = (x, y)$ for all x and y. But then this means that $(Mx)^T J(My) = x^T Jy$ for all x and y where $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = J$. This holds if $x^T M^T JMy = x^T Jy$ so it holds always if $J = M^T JM$. The matrices of this M for which this holds is a group as it contains the identity and is closed under inverses and products and matrix multiplication is associative. Also, we know that $Det(M) = \pm 1$ since the determinants have to agree. We will restrict this group to those with determinant +1. This is called the **Lorentz group**.

Note that $e_0^T M^T JMe_0 = e_0 Je_0 = 1$ and we can use this to deduce that $M_{00}^2 - M_{10}^2 = 1, M_{01}^2 - M_{11}^2 = -1$ since we have decided to start zero indexing things now just because we figured that would be funny. We can then write this as $\begin{pmatrix} \pm\cosh(\theta) & \pm\sinh(\theta) \\ \pm\sinh(\theta) & \pm\cosh(\theta) \end{pmatrix}$. We will restrict the cosh terms to be positive and then the sinh terms can be any sign depending on the sign of θ, so we can write $\begin{pmatrix} \cosh(\theta) & \sinh(\theta) \\ \sinh(\theta) & \cosh(\theta) \end{pmatrix}$ which we can now see forms a group since it is closed since multiplying them just adds the θs, so it is essentiallt the group of real numbers under addition (ie it is isomorphic). With the norms from the normal inner product, we got circles when we kept it constant. Now we can see that we will get hyperbolae when we do this. $x^T Jx = x_0^2 - x_1^2$ and we keep $x_0^2 - x_1^2 = c$ so we get a hyperbola. We can rewrite $M(\theta) = \frac{1}{\sqrt{1-\tanh^2\theta}} \begin{pmatrix} 1 & \tanh(\theta) \\ \tanh(\theta) & 1 \end{pmatrix}$. Define $v \coloneqq \tanh(\theta)$ and $t \coloneqq x_0, x \coloneqq x_1$. We can interpret v as the factor of the speed of light we are going at, and this has to be between -1 and 1. We can interpret $a' = Ma$ for a=(t,x) as:

$$t' = \frac{1}{\sqrt{1-v^2}}(t + vx), x' = \frac{1}{\sqrt{1-v^2}}(x + vt)$$

See level 8.5 to see why we are doing this: There is a physical interpretation for the idea that this gives relativistic effects.