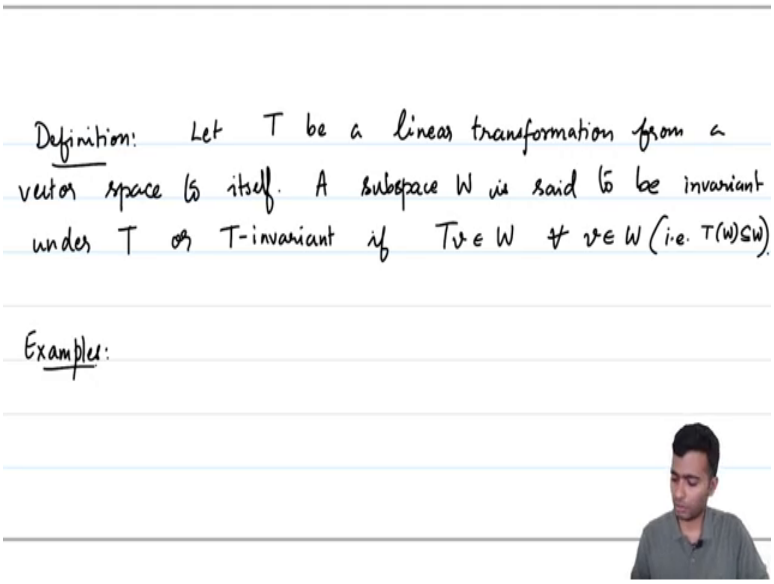
Linear Algebra
Professor Pranav Haridas
Kerala School of Mathematics, Kozhikode
Lecture 8.4
Invariant subspaces

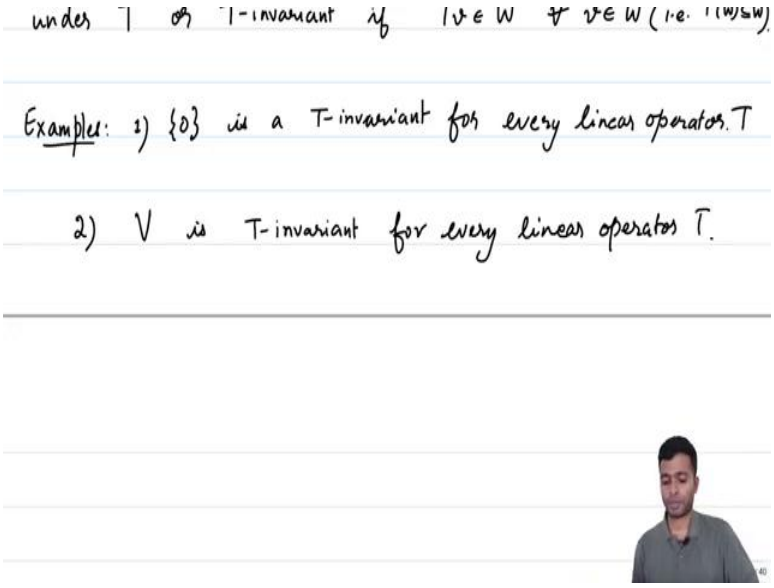
So, let us next discuss the notion of subspaces which are invariant under a given linear transformation. We will use that notion to prove a very celebrated theorem called the Cayley-Hamilton theorem. So, let us begin by defining the notion of a subspace which is invariant under a given linear transformation.

(Refer Slide Time: 0:40)



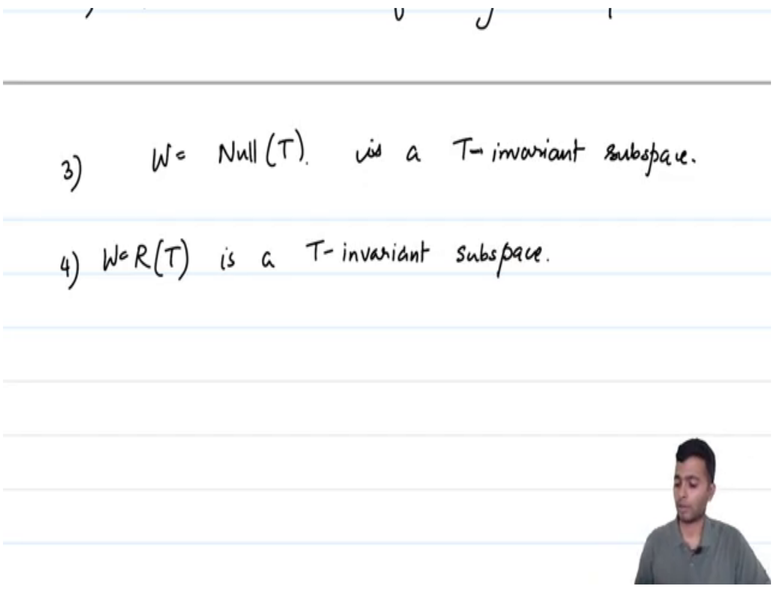
So, definition, so let T be a linear operator, let T be a linear transformation which is same as a linear operator from vector space to itself. We say that subspace W is invariant under T if it is preserved by T or in other words, if every vector of W is sent to W itself, then it is called as a subspace which is invariant under T. So, a subspace W is said to be invariant under T, or T invariant both terminologies are used if T v belongs to W for all v in capital W. So, the compact way of saying this is that T of W is contained in W, so let us look at a few examples.

(Refer Slide Time: 2:05)



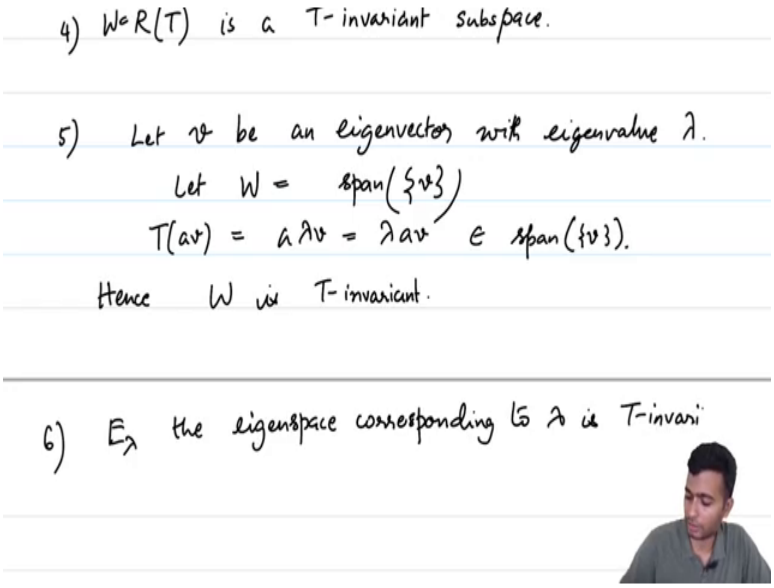
The first example we would like to consider is always a 0 vector space. So, 0 is a T invariant subspace, 0 vector subspace is T invariant for every linear operator T. So, notice that the invariance, the property of invariance is dependent on T. If T changes, if you take a subspace of say R n, which is invariant under one operator, it need not be invariant under a different operator. However, 0 vector space or 0 subspace is a subspace which T invariant under every linear operator. So, the entire vector space V is the invariant for every linear operator T. So, what more?

(Refer Slide Time: 3:18)



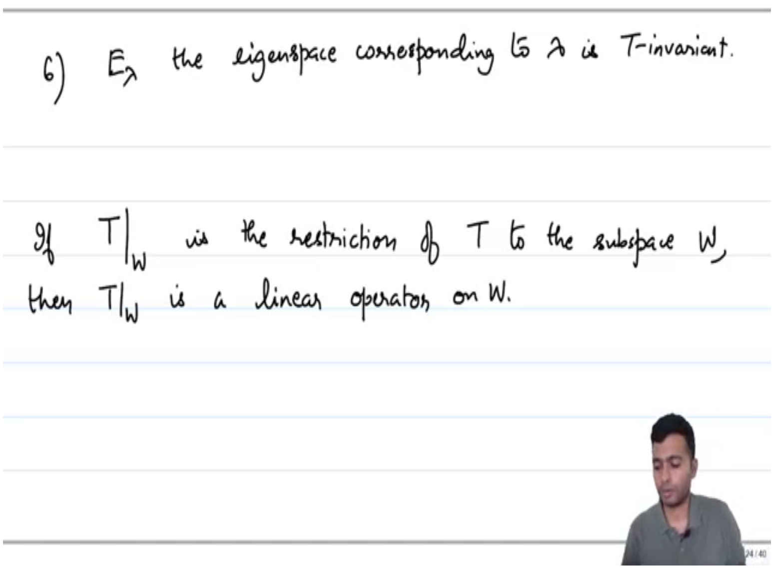
So, given any linear operator, consider W to be the null space of T, then notice that T v is equal to 0 for every v in W, and 0 in particular belongs to W and therefore, it is trivially a invariant subspace in a similar W equal to R of T, which is the range of T is a T invariant of subspace.

(Refer Slide Time: 4:11)



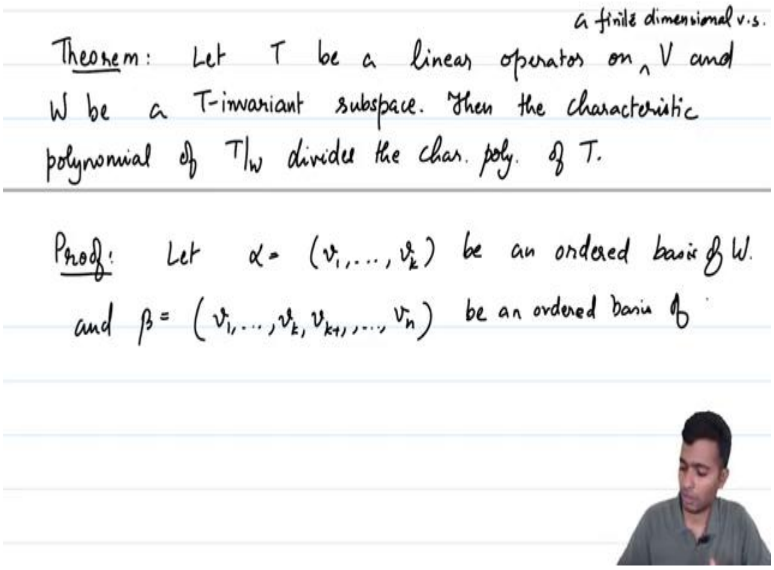
So, more examples. So, let Lambda be an eigenvalue or let me put it this way, let v be an eigenvector with eigenvalue Lambda and then let w be the span of the set consisting of v, s is v say span of 1 vector. So, this consists of all vectors of the type A times V, and what is going to be T of A times V by linearity this is a Lambda v which is equal to Lambda times a v, which you notice again belongs to the span of v itself, so hence, W is the T invariant that we could have set more, said that E Lambda, the Eigen space corresponding to Lambda is T invariant.

(Refer Slide Time: 5:43)



Same argument given in example 5 will tell us that E Lambda is also T invariant. So, one reason why we should be considering T invariant subspaces is because you take the linear operator T and restrict it to the subspace W. And we can think of it as a linear operator on W. So, if T restricted to W is the restriction of T to the subspace W, then T restricted to W is a linear operator on W.

(Refer Slide Time: 6:46)



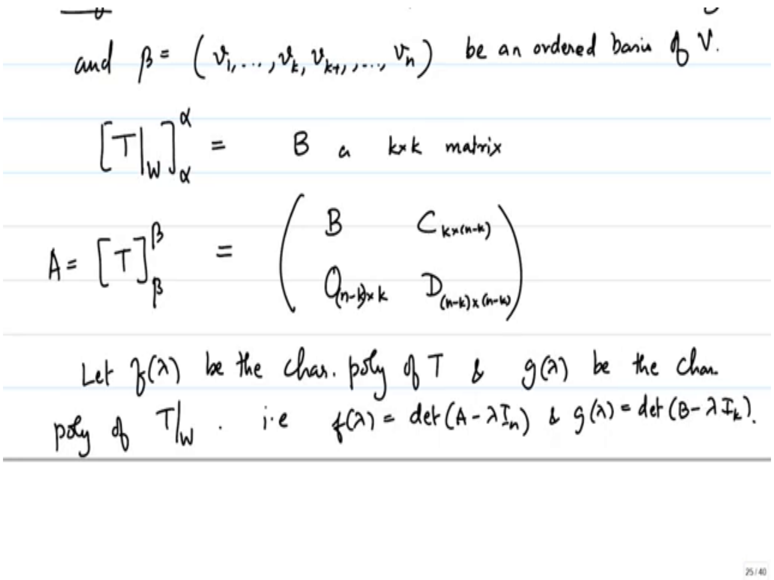
So, T restricted to W can be studied as an independent linear operator, which is operating on the subspace W. So, let us prove a theorem to indicate that this is quite useful. So, let T be a Linear Operator on V, and W be a T invariant subspace, then the characteristic polynomial of T restricted to W divides the characteristic polynomial.

So, let me just write Char for characteristic and poly for polynomial of T. So, let us look at a proof, so this theorem tells us that by studying the restriction of the linear operator to an invariant subspace we get some information about the characteristic polynomial, we get a hold of at least one factor of the characteristic polynomial.

Let us look at a proof of this, so let us start with basis for our subspace W. So, let alpha equal to v 1 to v k be an ordered basis of W. Now, this is linearly independent set, this is a linearly independent set which is sitting in capital V and hence it will be sitting inside a basis. So, extend this to a basis and Beta equal to v 1 to v k, v k plus 1, to v n.

So, let me add at this point that even though I am not writing it, we will consider only finite dimensional vector spaces. So, let me write here that it is a linear operator on finite dimensional, we are indeed talking about the characteristic polynomials. Let me write v dot s for vector space but nevertheless let me put it in the criterion. So, Beta we obtained by extending our basis alpha, Beta be a basis of V Beta be a basis of again ordered basis of V.

(Refer Slide Time: 9:54)



Let us look at T W with respect to Alpha that will be let us say that is some matrix B, then let us now look at what is T with respect to Beta. I leave it to you to check that this will be a matrix B which is of course, k cross k matrix. And there will be a 0 matrix consisting of n minus k cross k vector elements and then there is a C and that is a D. C is k cross n minus k and this is going to be the n minus k cross n minus k, this is matrix of T with respect to Beta will be here.

So, to talk about the characteristic polynomial, we have to look at the determinant of so let f of Lambda be the characteristic polynomial of T and G of Lambda be the characteristic polynomial of T restricted to W. So, let us call these things this as A. So, g of Lambda or rather f of Lambda is equal to determinant of A minus Lambda I n and g of Lambda is equal to determinant of B minus Lambda I k.

(Refer Slide Time: 11:53)

Let $f(\lambda)$ be the char. poly of T & $g(\lambda)$ be the char. poly of $T|_W$. i.e $f(\lambda) = \det(A-\lambda I_n)$ & $g(\lambda) = \det(B-\lambda I_k)$.

$$f(\lambda) = \det(A-\lambda I_n) = \det\begin{pmatrix} B-\lambda I_k & C \\ 0 & D-\lambda I_{n-k} \end{pmatrix}$$

$$= \det(B-\lambda I_k)\,\det(D-\lambda I_{n-k}).$$

Let us analyse what is determinant of A minus Lambda I n will be, this is equal to this is what f Lambda is. This is equal to determinant of, let us look at what A is, A has four blocks, the bottom block has a 0, left bottom block and the right top block will not be affected at all by subtracting Lambda I, so this is just going to be B minus Lambda I k, that will be a C that will be a 0 here and there will be T minus Lambda I n minus k. And we know exactly what this determinant is by 1 of the exercises earlier, this is going to be equal to the determinant of B minus Lambda I k times the determinant of D minus Lambda I n - k.

(Refer Slide Time: 12:48)

& hence $g(\lambda) \mid f(\lambda)$. —— ■

Another example of a T-invariant subspace

Let $v \in V$ and define

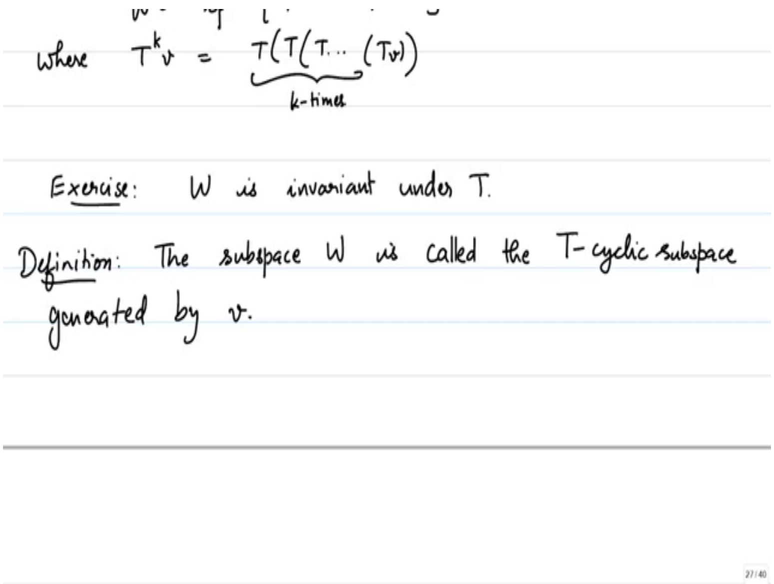
$$W = \text{span}\{v, Tv, T^2v, \dots\}$$

where $T^k v = \underbrace{T(T(T\dots}_{k\text{-times}}(Tv))$

But what is this? This is equal to g of Lambda and this is going to be some polynomial P of Lambda, and hence g of Lambda divides f of Lambda and we are done with the proof. So, if we have a subspace w which is invariant under T and if we restrict our attention to restrict our T to W and look at the characteristic polynomial there, it divides the characteristic polynomial of T itself. We will use this result and a few more other observations to very particular subspaces of V which are invariant under T. So, let us look at one more example another example.

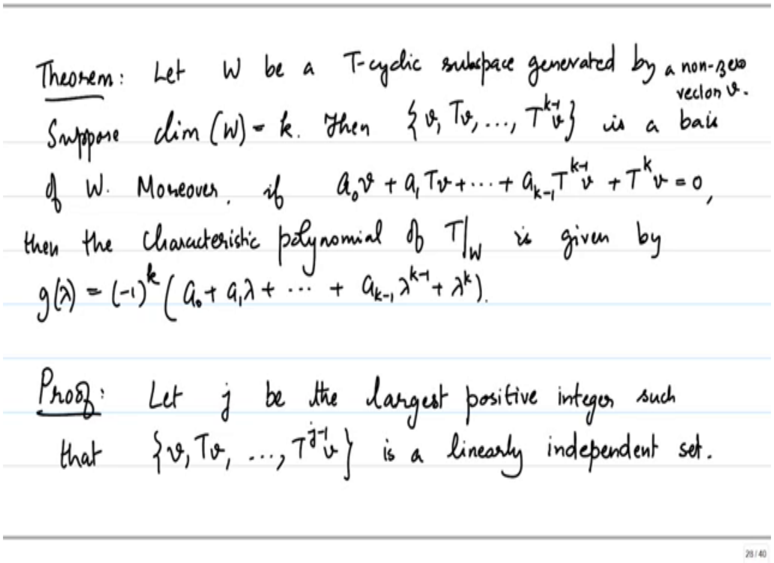
Example of T invariant subspace. So, let v be a vector in capital V, then let W, and define W to be the span of v, T v, T 2 v and so on. So, what is T 2 v where T k v is just T of T of T of T v, what is that? T is a linear operator from V to itself so, T v is a vector in capital V, so we can apply T to T v that will be T square v. T k v is similarly done k times. So, get hold of a vector v and look at all powers of T applied to V and then look at the span of this and I leave it as an exercise for you.

(Refer Slide Time: 15:21)



W is invariant under T and definition, the vector space the subspace W is called the T cyclic subspace generated by v, the subspace W is called the T cyclic subspace generated by small v. So, what can we say about this particular T cyclic sub subspace? This particular T invariant subspace, this T invariant subspace has some very nice properties, the first among them is that we have a very good control over what the basis of such a subspace is. And secondly, we know how the characteristic polynomial looks like when T is restricted to W.

(Refer Slide Time: 16:31)



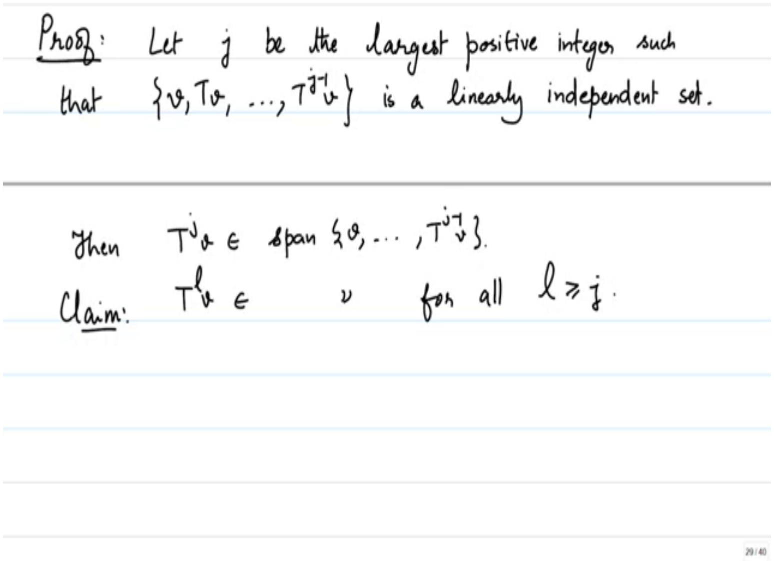
So, let us prove a theorem which captures all the information which I just said. So, let W be a T cyclic subspace generated by small v. Suppose, dimension of W is equal to say k, then v, T v up to T k minus 1 v is a basis of W, now exactly what the basis of W will look like. Not just that. Moreover, if so t to the power k v will be in the span of T v up to T k minus 1 v, and therefore we can write it as a linear combination.

So, if a 0 v plus a 1 T v plus a k minus 1 T k minus 1 v plus T k v is equal to 0, suppose this is the 0 vector, suppose this is the linear combination that gives us minus of T k v, then the characteristic polynomial of T restricted to W is given by minus 1 to the power k times a 0 plus a 1 Lambda which is given by g of Lambda, which is equal to minus 1 to the power k times a 0 plus a 1 Lambda plus a k minus 1 Lambda to the power k minus 1 plus Lambda to the power k.

Let us give a proof, so this is a new idea maybe so let us see. So, let j be the smallest positive integer such that v, T v, T square v up to T to the power j minus 1 V is linearly independent. So, let j be the largest positive integer, smallest would just be j equal to 0 so that is not useful, J be the largest positive integer such that v T v up to T to the power j minus 1 v, this is a linearly independent set.

So, let us put one added assumption in the theorem that v is by a non-zero vector v to just avoid unnecessary complications, non-zero vector v. When v is 0 there is nothing interesting that happens because then it will just be the zero vector.

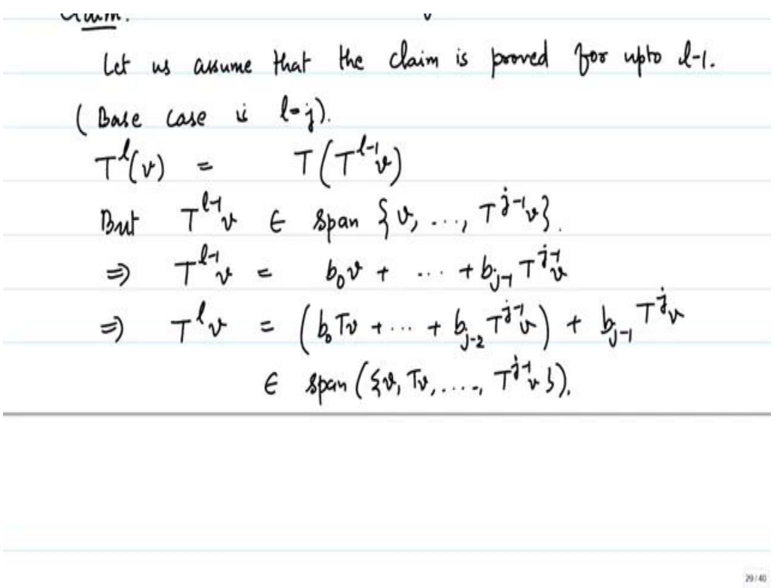
(Refer Slide Time: 20:45)



So, let j be the largest positive integer such that v, T v, up to the power j minus 1 v is linearly independent. Notice that for any value less than j, this is going to be again linearly independent because it is a subset of linear independent set. And also note that then T to the power j v belongs to the span of v T v up to T to the power j minus 1 v.

So, remember that T to the power j v means that T applied to T v apply so T applied to T applied to T applied to finally T v j minus 1 time. So, that is what T to the power j v means and that is in the span of v, T v up to T to the power j minus 1. So, let us assume that so we will prove that. So, the claim is that T l v belongs to the span of v, T v up to T to the power j minus 1 V for all l greater than or equal to j.

(Refer Slide Time: 22:10)



So, we will prove this by an induction argument. So, we already know that T to the power j v belongs to the span. So, proof let us prove by induction, so let us assume that the claim is proved for up to l minus 1, we will prove that T to the power l v also belongs to the base case where l is equal to j has been here the base case remember is l is equal to j.

Let us now prove that T to the power l v also belongs to the span. But then T to the power l v is nothing but T acting on T to the power l minus 1 v by the very definition of T to the power l v. And we know that T to the power l minus 1 v belongs to the span of, I would have given it some name so that I did not have to write it down so many times, but that is okay.

But that implies, t to the power l minus 1 v is equal to maybe b 1 v plus or b 0 v plus b j minus 1 T j minus 1 v. And hence, T to the power l v will just be equal to b 0 T v plus b j minus 2 t j minus 1 v plus b j minus 1 times T j of v, but T j of v we already noted belongs to the span of v, T v up to T to the power j minus 1 v, and therefore, that will be a linear combination of these vectors. Therefore, this belongs to the span of v, T v up to T to the j minus 1 v and therefore we are done.

(Refer Slide Time: 24:27)

$$\Rightarrow \quad T^{l}v = (b_0 Tv + \cdots + b_{j-2} T^{j-1} v) + b_{j-1} T^{j} v$$

$$\in span(\{v, Tv, \ldots, T^{j-1}v\}).$$

$$\text{i.e} \quad T^{l}v \in span(v, \ldots, T^{j-1}v) \quad \forall\, l.$$

$$\Rightarrow \quad W \subset span(v, \ldots, T^{j-1}v) \subseteq W$$

$$span(v, \ldots, T^{j-1}v) = W.$$

30/40

$$\text{Span}(v_1, \ldots, T^{j-1}v) = W$$

Hence $\{v_1, \ldots, T^{j-1}v\}$ is a basis of W.

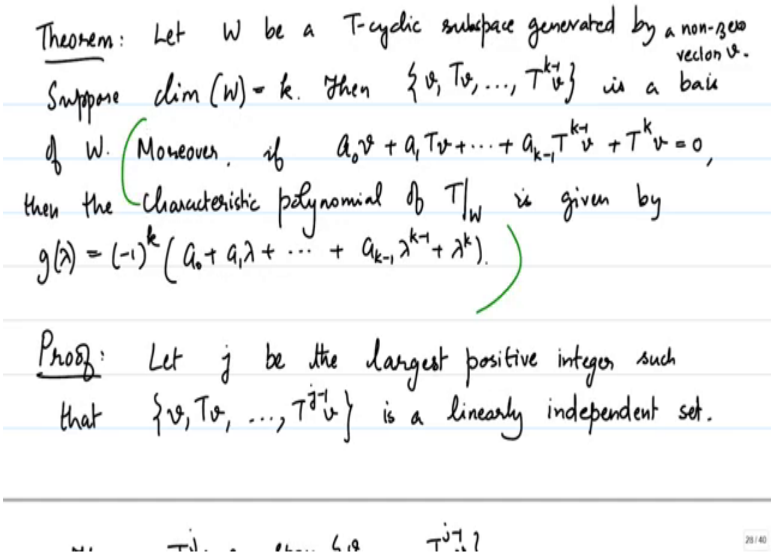
Since $\dim(W) = k$, we have $j = k$.

30 / 40

But what does that mean? This means that T to the power l v belongs to, I should have really written a notation for it, this for all l, which implies that what is the name given for T cyclic subgroup is called W that means that w is contained in span of v 1 to T v j minus 1 v, but this is again contained in W, because each of these are in w and therefore this one should be contained in that. Therefore, span of v 1 to T j minus 1 v is equal to W, which has okay, we will come to that. So, what do we know now? We know that we v, T v up to T to the power j minus 1 v is both a spanning set and linearly independent and therefore it is a basis, hence v to T to the j minus 1 v is a basis of W.

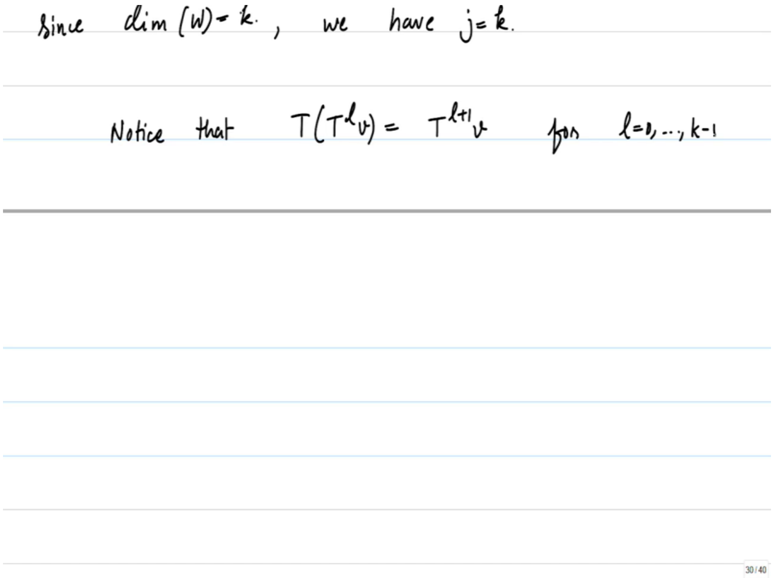
What do we know about the basis of W? Know that it has k elements, since dimension of W is equal to k, we have j is equal to k. And we are done because recall what we were trying to prove.

(Refer Slide Time: 26:05)



We were trying to prove that v, T v up to T to the power k minus 1 v is the basis of W, and that is precisely what we have ended up proving. We are not done, we have to prove the second part, this part we have to show or we have to get explicitly how the characteristic polynomial will look like. So, to do that, again, let us look at what we know about v, T v and so on.

(Refer Slide Time: 26:43)



Let $-T^k v = a_0 v + a_1 T v + \cdots + a_{k-1} T^{k-1} v.$

Then for $\beta = (v, Tv, \ldots, T^{k-1} v)$,

$$[T|_W]_\beta^\beta = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & & \vdots & -a_1 \\ 0 & 1 & & \vdots & \vdots \\ \vdots & 0 & & & \vdots \\ & \vdots & & 0 & \\ 0 & 0 & & 1 & -a_{k-1} \end{pmatrix}$$

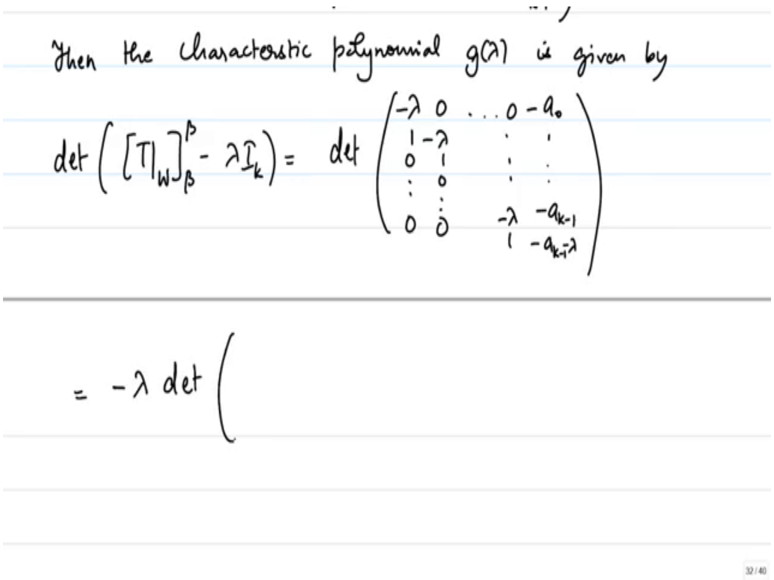
31/40

Notice that T of T l v is equal to T l plus 1 v for l is equal to 0 to so T to the power 0 by convention is the identity map, so this will be up to k minus 1. In fact, it is for all so, let me not write it this way, let me just say that let T to the power k v be equal to a 0 or minus let me put it that way, a 0 v plus a 1 T v plus up to a k minus 1 T k minus 1 v. Suppose, this is the linear combination which gives us minus T to the power k v, then a 0 v plus then let us see then what happens, what is the matrix of T going to look like? We are interested in the matrix of T restricted to W.

And we know explicitly how this matrix will look like with respect to Beta, where Beta I should have given it from the beginning, so let me give it a name now, for Beta equal to v, T v, T to the power k minus 1 v. Suppose Beta is this, let us look at what is T restricted to W Beta Beta. So, what is T v? T v is the second vector in the order basis.

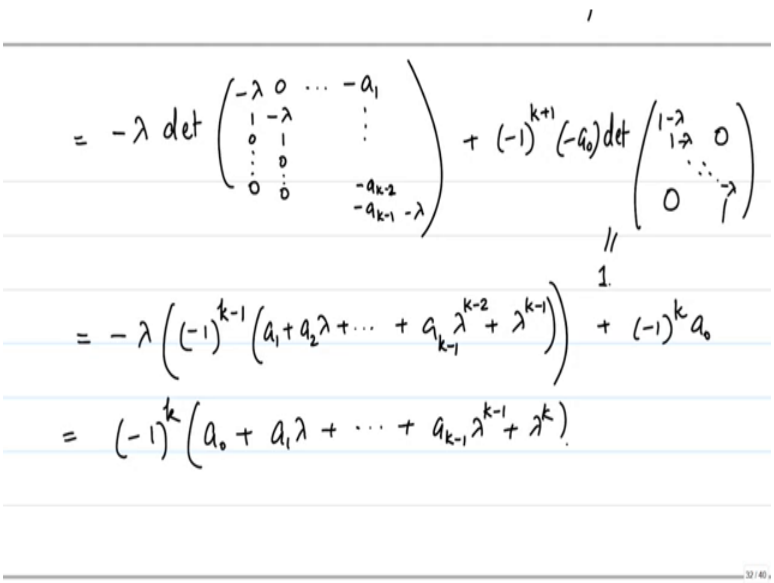
So, if you notice, we just write it for you, the first column will be the image of v, which is going to be 0 1 0 up to 0 all the way down. And similarly, T v is going T 2 v which will be a 0 here and so on. And T k minus 2 will be sent to, this is the k minus 1 basis vector or whereas T k minus 2 of v being sent. T k v is where we will have to worry about, T k v will just be sent to minus of a 0, minus of a 1, minus of a k minus 1. This is exactly what the matrix of T restricted to W will be with respect to Beta.

(Refer Slide Time: 29:58)



So, let us see what is then the characteristic polynomial. Let us call it g, g Lambda is given by determinant of this matrix minus Lambda times I k, which is just going to be equal to, let me write down the matrix explicitly. So, I know what the matrix is from here, this will be minus of Lambda 1, 0 all the way down 0 minus Lambda 1 0 all the way down minus Lambda and a 1 and then there will be minus a 0 up to a k minus 2 minus of a k minus 1 minus Lambda. This is precisely what our expression will look like. So, we will prove that the, so let us do a cofactor expansion along the first row, this is just going to be equal to minus of Lambda times the determinant of.

(Refer Slide Time: 31:09)



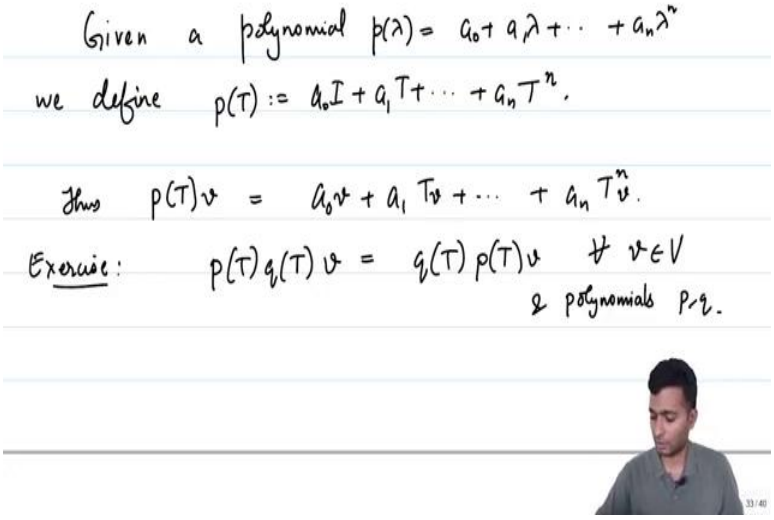
So, the first row and the first column is deleted and what remains is this, and there will be a 0 minus Lambda 1 0 and so on. And finally there will be a minus of a 1 to minus of a k minus 2 and then there is a k minus k minus Lambda. And how about the last term so there are k vectors, so this is just going to be minus 1 to the power k plus 1 times minus of a 0 into the determinant of by deleting the first row and the last column, which will just turn out to be ones in the diagonal, 0 below minus Lambda in the off diagonal and then 0 here. And the determinant here will just be equal to 1.

So, this is equal to V, the determinant of upper triangular matrix will just turn out to be the product of its diagonal entries. Then what do we do about the first term here? We will use an induction argument which I will allow you to complete. By an induction argument, this will be equal to now in this case, it is going to be minus 1 to the power k minus 1 times let us see.

This one starts from a 1, so this is going to be a 1 plus a 2 Lambda plus a k minus k does not make sense, a k minus 1, Lambda to the power k minus 2 plus Lambda to the power k minus 1. And the term here will just be equal to minus of 1 to the power k plus 2, which is the same as minus of 1 to the power k times a 0.

And this is nothing but the minus will add up, this will be a minus of 1 to the power k, the a 0 will come here and the Lambda multiplied will give you a 1 Lambda goes up to a k minus 1 Lambda to the power k minus 1 plus Lambda to the power k and we are done. So, when our subspace W is the T invariant subspace generated by a vector v, we know exactly what the characteristic polynomial of T restricted to W will look like. So, now we have all the ingredients that are needed to prove the celebrated Cayley-Hamilton theorem.

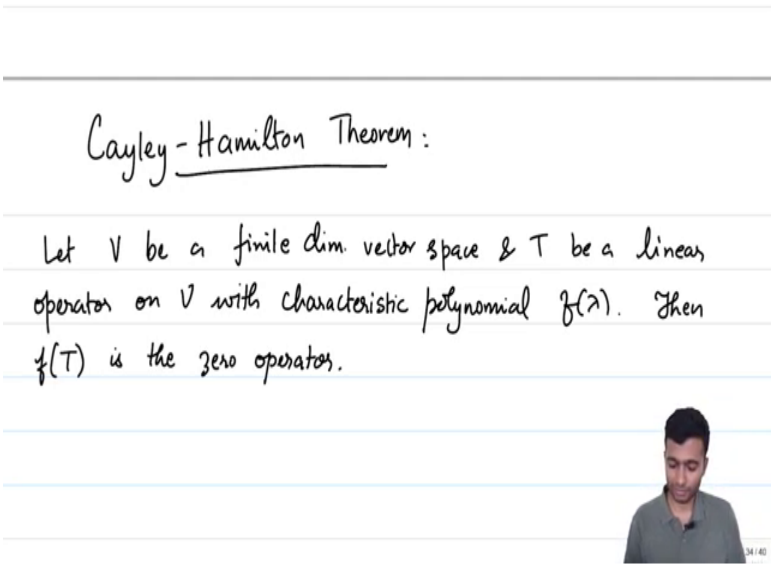
(Refer Slide Time: 34:29)



So to do that, let me just develop a few notations. So, given a polynomial say P of Lambda, which is equal to a 0 plus a 1 Lambda plus a n Lambda to the power n, we define P of T to be equal to is a 0 i plus a 1 T plus a n T to the power n. And therefore thus, what is the meaning of P T acting on a vector V, this is just going to be equal to a 0 I acting on v which is equal to v plus a 1 acting on T v plus up to a n acting on T to the power n v.

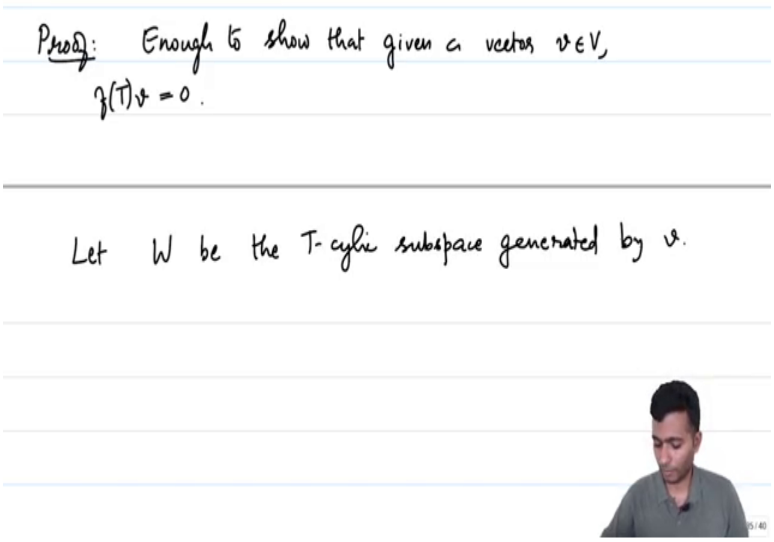
So, this is going to be a linear combination of v T v up to T to the power n v. I will leave it as an exercise for you to check that p T q T of v is equal to q T, p T of v for all v in capital V and polynomials p, comma q. It is quite straightforward I will leave it as an exercise for you to take that.

(Refer Slide Time: 36:16)



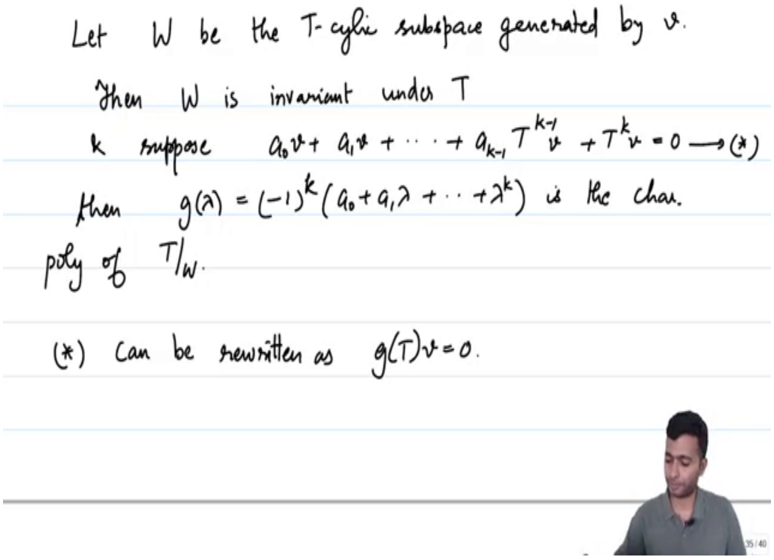
Now, let us state and prove the Cayley-Hamilton theorem. The Cayley-Hamilton theorem broadly tells us that any linear operator satisfies its own characteristic polynomial. So, let me write it down in words, so let V be a finite dimensional vector space and T be a linear operator on V with characteristic polynomial f of Lambda, then f of T which we have defined a few minutes back is the 0 operator. It is the linear transformation, which sends every vector V to 0. Informally, we say that T satisfies its characteristic polynomial. So, let us give a proof of this.

(Refer Slide Time: 37:32)



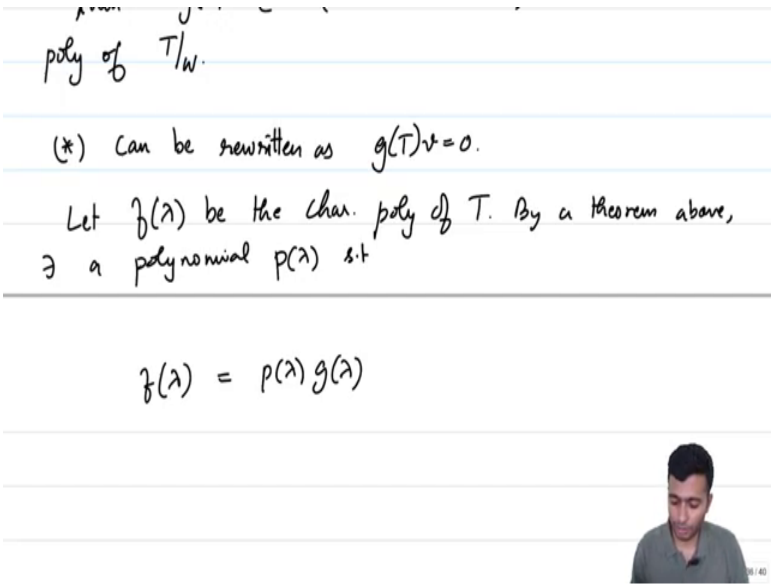
Enough to show that for given a vector v in capital V, f of T v is equal to 0. So, if we do that by choice our vector v is arbitrary, so this will be satisfied for every v. So, how do we go about doing this? So to do this, let W be the T cyclic subspace generated by v.

(Refer Slide Time: 38:45)



So, what do we know about the characteristic polynomial of T restricted to W? Then W is invariant, first observation is that W is invariant under T and suppose a 0 v plus a 1 v plus a k minus 1 T k minus 1 v plus T to the power k v is equal to 0, then g of Lambda equal to minus 1 to the power k times a 0 plus a 1 Lambda plus up to Lambda to the power k is the characteristic polynomial T restricted to W. But what do we know about g of T v. So, let me write it as star here, star can be rewritten as g of T acting on V being equal to 0 vector.

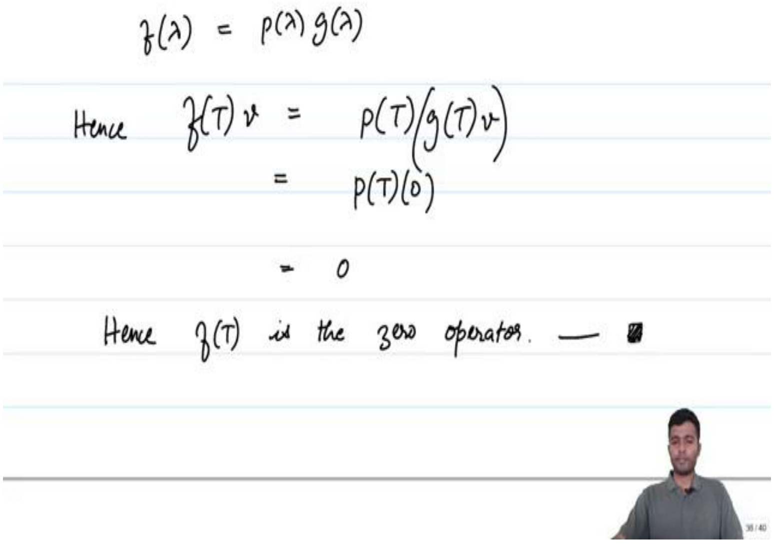
(Refer Slide Time: 40:17)



We also proved that if W is T invariance subspace, then the characteristic polynomial of T restricted to W divides the characteristic polynomial of T. So, let f of Lambda be the characteristic polynomial of T, then by a theorem above where exist a polynomial say P of

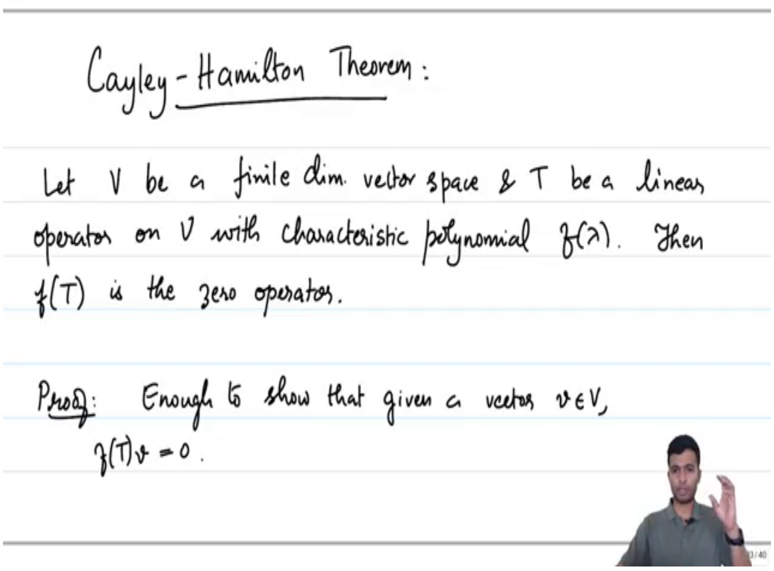
Lambda such that f of Lambda is equal to p of Lambda times g of Lambda, where g of Lambda is the characteristic polynomial of T restricted to W.

(Refer Slide Time: 41:22)



And hence, what is going to be f of T v. That is just going to be equal to p of T times g of T of v, but this is already equal to 0, this is equal to p of T, the operator acting on the 0 vector, but any operator takes any linear transformation takes 0 to the 0, therefore this is equal to 0. And hence f of T is the 0 operator that is the completion of our proof.

(Refer Slide Time: 42:10)



So, I would like to conclude by observing that if we stated Cayley-Hamilton theorem as the following, let a be an n cross n matrix and f of Lambda be the characteristic polynomial of a, then f of a is equal to 0, where a to the power n is the product of a with itself n times and f of a is just the polynomial expression which involves a to the power n, a to the power n minus 1 and up to identity.