Linear Algebra

Professor K.C Shivakumar
Department of Mathematics
Indian Institute of Technology, Madras

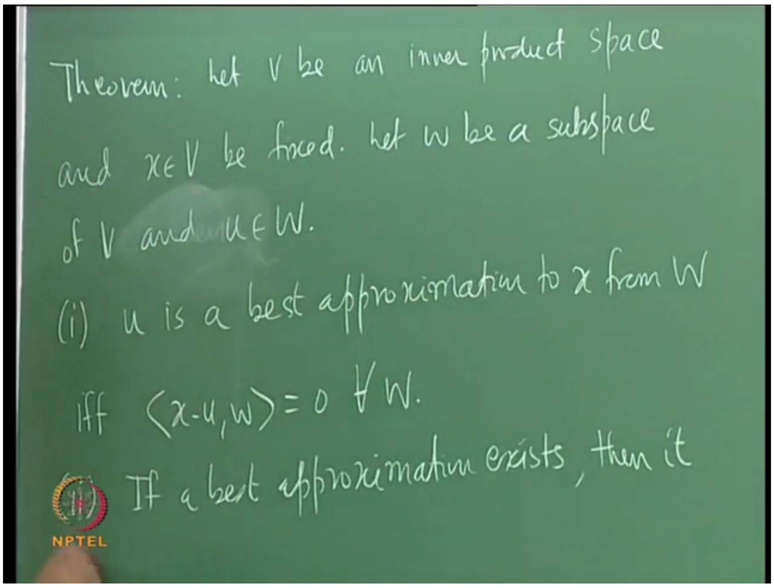
Module 12-Best Approximation

Lecture 43

Best Approximation: Least Squares Solutions

Ok so are discussing the problem of best approximation, I gave an example last time and stated a theorem lets look at a proof of the theorem. Let me write down the theorem once again so that we will keep on this side for reference.

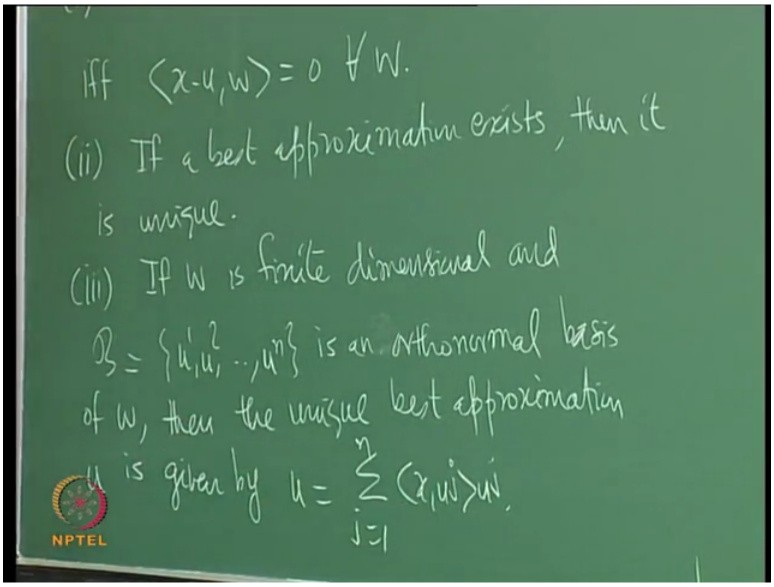
(Refer Slide Time: 00:37)



We have an inner product space and a finite dimensional subspace ok so lets take V be an inner product space and x is a fixed element. Let W be a subspace not an arbitrary subset be a subspace of V then we have the following. I think I also (())(1:29) belongs to W right, let W be a subspace of V and U be an element in W, is that fine ok. We are interested in knowing when this U is a best approximation to x from W.

When is U a best approximation to x from the subspace W ok, the first assumption is that U is a best approximation to x from W there is a(())(2:15) condition if and only if x minus U is orthogonal to the subspace W. Second condition, if a best approximation exists then it must be unique ok, then it is unique.

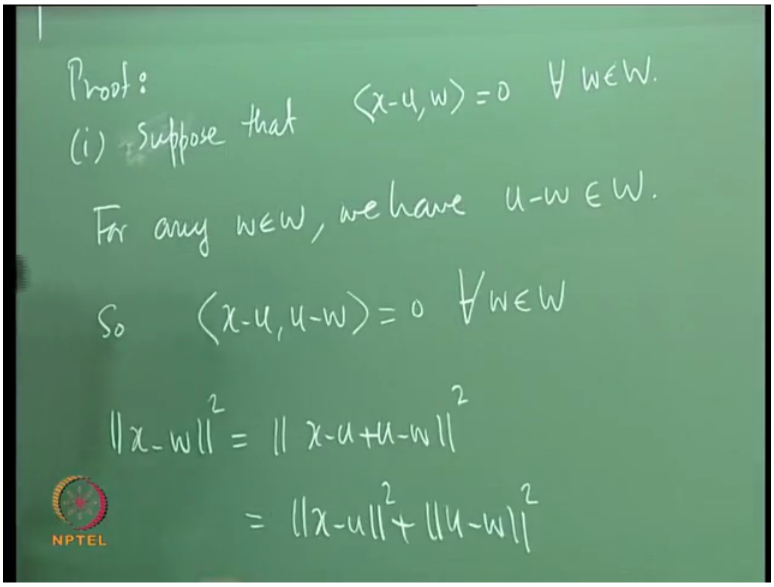
(Refer Slide Time: 02:53)



Third condition if W is finite dimensional and lets say I have an ordered orthonormal basis U1 U2 etc Un this is an orthonormal basis, then the unique best approximation U is given by this formula ok, x is fixed this U is greater to x by means of this coefficients. So if this U1 U2 etc Un is an orthonormal basis of W then this is the formula for the unique best approximation to x from W.

We have already seen an example, where this formula was used. Find the point on the plane which is closest to a vector 1 2 1.

(Refer Slide Time: 04:41)

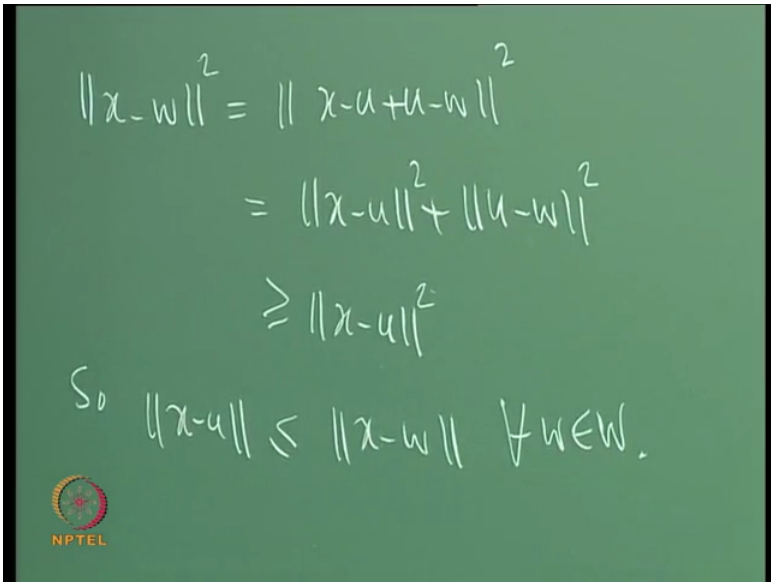


Ok proof, we need to show that this is a necessary (())(4:47) condition, a formula that could be used ok. Lets take lets prove this part first, suppose that x minus U is perpendicular to W we must show that U is a best approximation ok. Remember that the best approximation in the definition this U must come from W ok. So if you take any w in W it must follow that U minus W belongs to W, W is subspace and so this condition can be rewritten as x minus U, U minus W this must be zero for every W and W.

Remember U is fixed just as how x is fixed U I know is a vector that satisfy this equation U is a fixed vector in W that satisfies this equation, its only small w that is variable here, so this is true for all W and W I need to show that U is a best approximation I have this I will use Pythagoras theorem. Consider x minus W the whole square this is you subtract and add U x minus U plus U minus W the whole square, subtract and add now observe x minus U and U minus W are orthogonal because the previous equation and so by Pythagoras theorem.

Norm x plus y square is norm x square plus norm y square if x is perpendicular to y so this is norm x minus U square plus norm U minus W square but this is always greater than or equal to norm x minus U square because this is non-negative norm U minus W square is non-negative so this inequality holds and I can take the square root because this is non-negative this is non-negative.

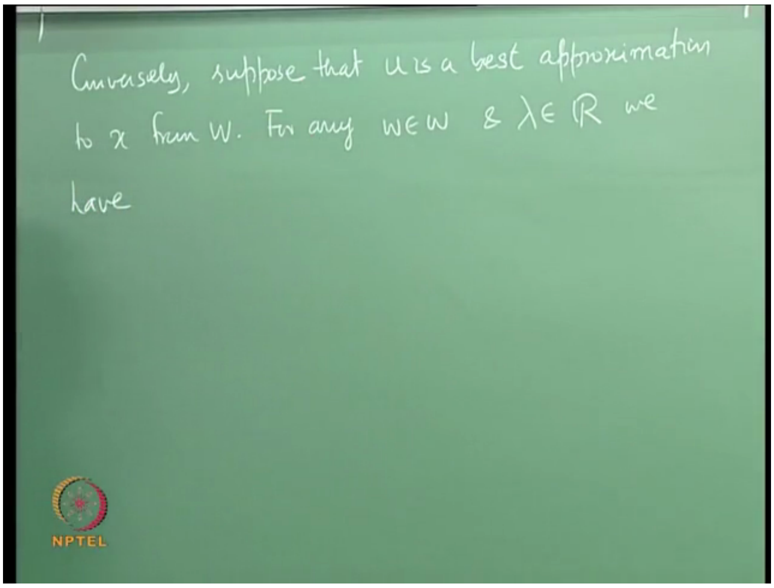
(Refer Slide Time: 07:03)



So it follows that norm x minus U is less and or equal to norm x minus W this is true for all W and W and this is precisely what we mean by saying that U is the best approximation.

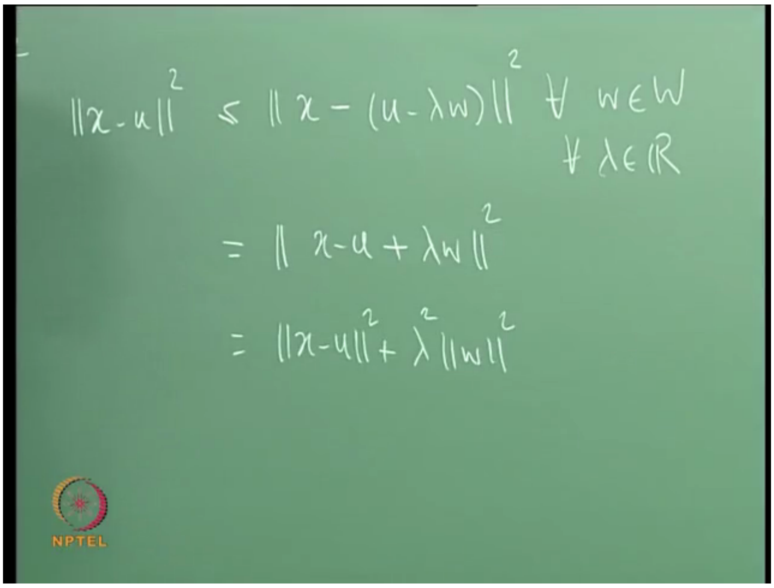
Ok converse, suppose U is the best approximation we must show that this is zero ok.

(Refer Slide Time: 07:35)



So conversely ok conversely suppose that U is a best approximation to x from W we must show that U satisfies a condition that x minus U is perpendicular W ok. For any W in W and lambda in R, see V is an inner product space so it might be a complex inner product space and restricted my attention to scalars that are real for any W in W and lambda in R what follows is that remember that we have assumed that U is a best approximation.

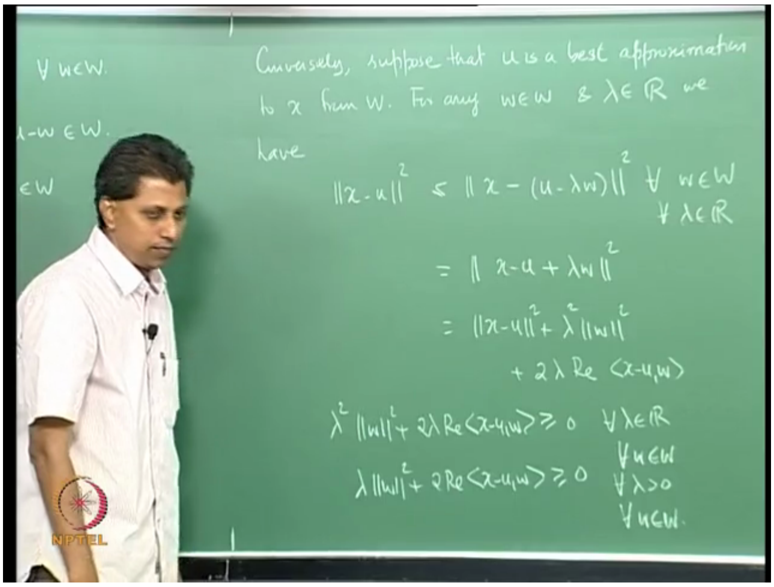
(Refer Slide Time: 08:39)



So if you look at norm of x minus U that must be less and or equal to norm of x minus any vector that belongs to W, in particular U minus lambda W this is true for all lambda sorry for all W in W and for all lambda in R ok then I square I will expand this using the inner product see what I get. This is equal to inner product I can rewrite this as norm of x minus U plus lambda W the whole square and then do the usual expansion. Norm of x minus U square is one of the terms plus this will go with (lambda) this will be lambda lambda bar but lambda is real so lambda square norm W square then there will be two terms which can be written as two times plus two times lambda is a scalar look at real part of x minus U with W.

This is what one what get after expanding is (right answer) lambda is real so this is taken out. So if you compare with the left hand side this x minus U the whole square can be cancel.

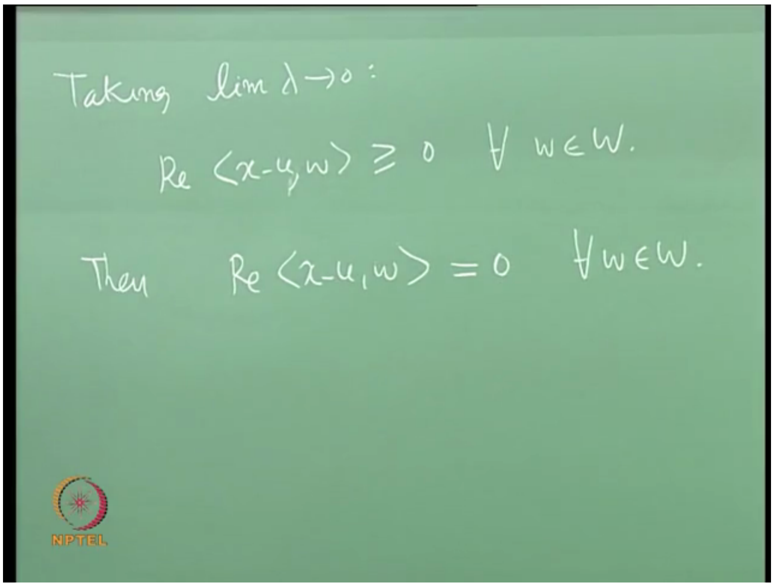
(Refer Slide Time: 10:08)



So what I get is lambda square norm W square plus 2 lambda real part of x minus U W this is greater than or equal to zero, now this is zero for all lambda element of R and for all W in W. in particular for lambda positive this is true so I will divide take lambda positive divide by lambda to get lambda times norm W square plus two times real part of x minus U W, this is greater than or equal to zero for all lambda positive this time and for all W in W.

Now why did I divide by lambda and made my did I made my did I make the assumption lambda positive in order to retain the inequality ok. Now take limit as lambda goes to zero for instance take lambda to be 1 by n the sequence 1 by N then you take the limit this is gone, what it means is that sorry I will have total lambda yeah fine. So what it means is real part x minus U W is greater than or equal to zero.

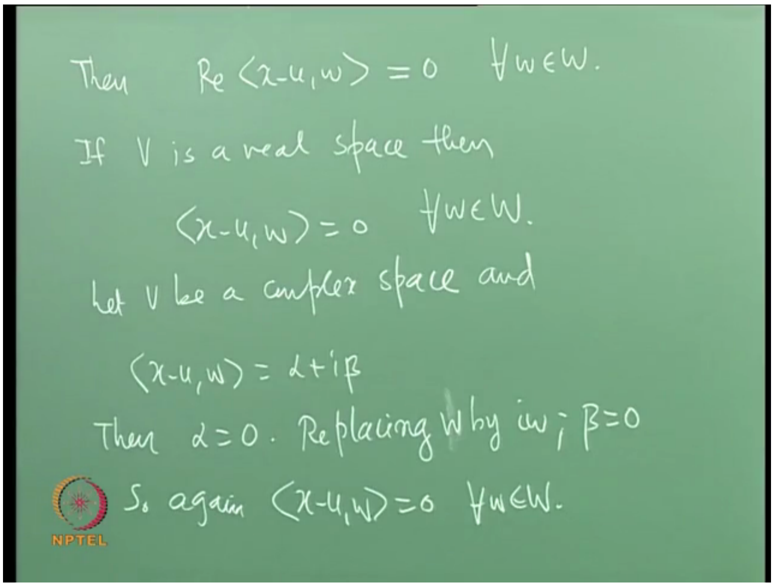
(Refer Slide Time: 11:26)



So taking limit lambda goes to zero we get real part of x minus U W this is greater than or equal to zero for all see lambda was a variable I have taken that as a sequence converging to zero W is only variable now for all W and W real part of x minus U W must be greater than or equal to zero.

W is a subspace so I can replace small w by minus W to get real part of x minus U W to be less than or equal to zero combine you get real part of x minus U W to be zero for all W and W ok that is I have a applied I have replaced W by minus W into this and that is a valid operation because capital W is a subspace so I get this equation plus consider two cases it is real inner product space.

(Refer Slide Time: 12:39)

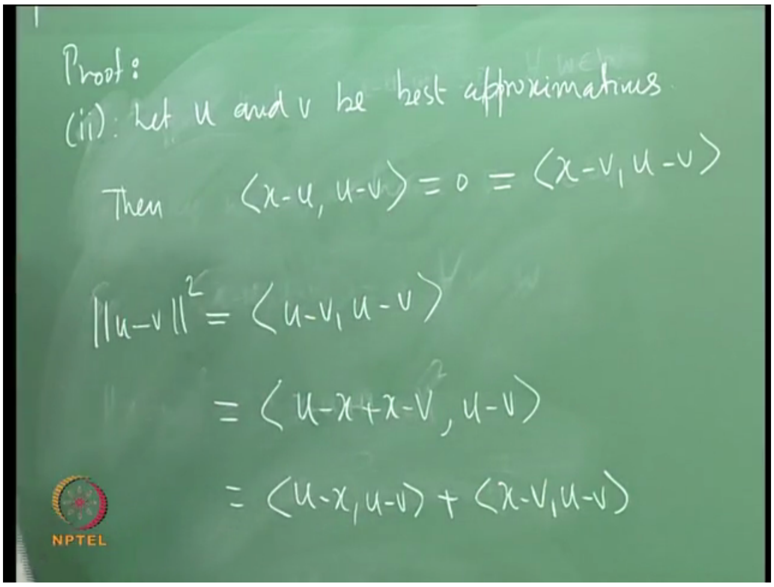


If V is a real inner product space then we have proved what we wanted, x minus U is perpendicular to the subspace W if U is a best approximation then x minus U is perpendicular to the subspace W.

If V is a complex inner product space if V is a complex space this real of this is zero so lets say if let V be a complex space and lets take x minus U, W to be alpha plus I beta then what we have proved just now is that then alpha is zero, real part of x minus U W 0 we have proved. Now this equation is true for all W so I will replace W by I W it is a complex space so scalar can be taken from the complex numbers, replace W by I W since this is true for all W you replaced W by I W then it will mean minus beta is zero. Replacing I by I W it comes in the second argument so when it goes out it goes with an I bar, I bar into I is a minus 1 or a plus 1, I bar is minus I so that is plus 1.

So replacing I W we get beta zero, real part of x minus U I W is zero but real part of x minus I W is just beta. So in any case this is zero. So once again just check this steps ok if necessary replace W by I W to conclude that x (minus) the imaginary part is also zero. So that’s the first part necessary sufficient condition for U to be best approximation is that this condition holds ok. Yeah replace W by I W ok.

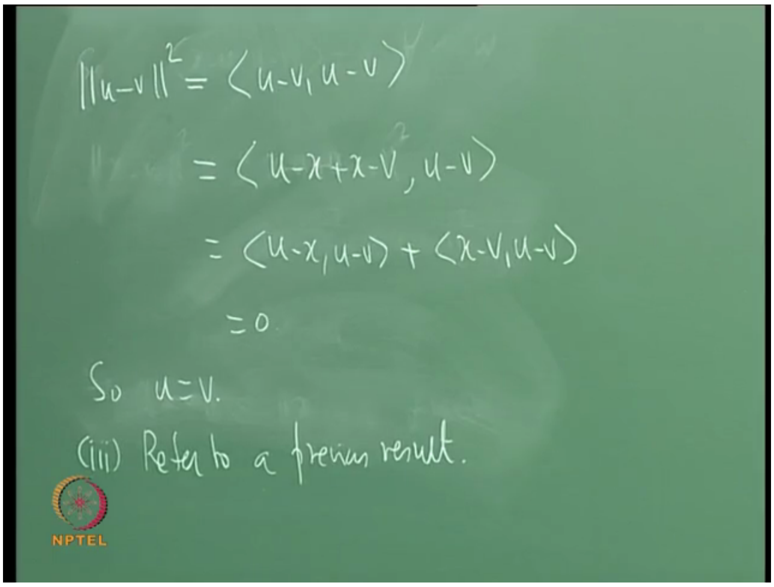
(Refer Slide Time: 15:35)



Second part is uniqueness, uniqueness typically to prove uniqueness you will assume that there are two best approximations and show that they are the same.

So let U and V be best approximations to x from W then from the first part x minus U, U minus V see U and V are best approximations so by definition they both come from W, so their difference is in W so this must be zero, x minus U, U minus V is zero similarly x minus V U minus V is zero by the first part U and V belong to W we have made use of that. I want to show U is equal to V consider norm U minus V square show that is it is zero norm U minus V square you want to show that this is zero then follows U is equal to V, you write it using inner product, U minus V U minus V this time keep the second term as it is, modify the first one. U minus x plus x minus V U minus V then that's U minus x U minus V plus x minus V U minus V both are zero.

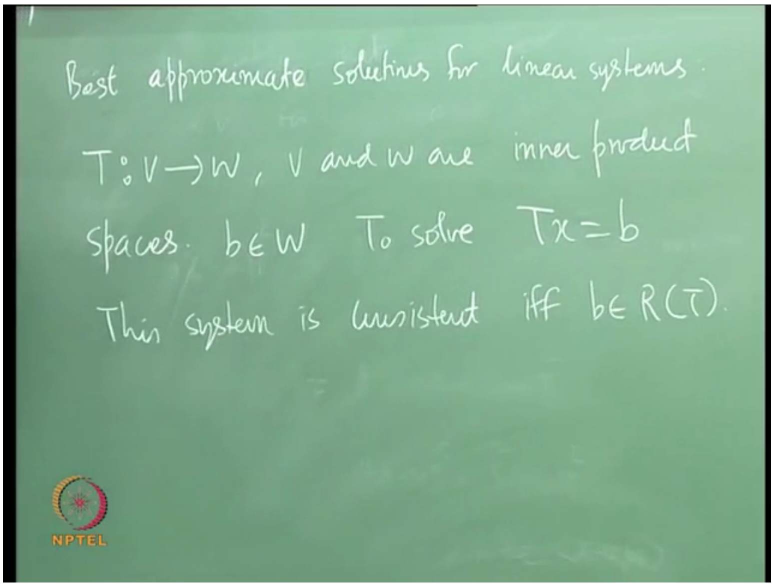
(Refer Slide Time: 17:06)



Ok, so U is equal to V if a best approximation exists then it must be unique. The last part gives you a formula in case W is finite dimensional. So prove the last part that this is the best approximation to z from W all you need to do is to show that x minus U is perpendicular to W then from the first part it follows that this is the unique best approximation ok. So just show x minus U is perpendicular to W that is what we will do but that has been done before, so the proof is over.You remember that in the very first theorem on best approximation this has been done this part was shown ok.

So let me just write for the third part refer to a previous result where we have shown that x minus U is perpendicular to W ok this is for best approximation from a subspace there is a related question which will seek best approximation solution for linear equations.

(Refer Slide Time: 18:45)

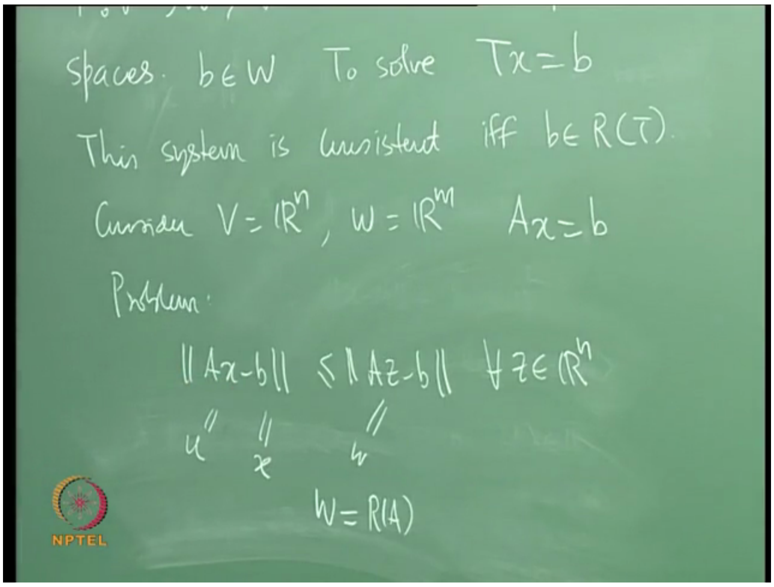


Best approximate solutions for linear systems of the form T x equal to B so let me pose the problem first, best approximate solutions for linear systems this is a more important problem than the bets approximation but using what we have developed just now this problem can be handled.

So what is the problem? T is a linear map between inner product spaces V and W are inner product spaces I have a fixed vector V from the (())(19:47) my question is how to solve T x equal to b more existential question, does this equation have a solution? This has a solution if and only if b belongs to range of T ok this system is consistent that is it has a solution if and only if b belongs to range of T but often b may not belong to range of T typically in statistical applications in experiments that (())(20:36) laboratory for example these equations will not be consistent but one still wants to solve these equations it is not consistent but can be atleast minimize norm T x minus b that is the question ok.

I will give you a examples in a real inner product space finite dimensional so I will confine my attention to the case when V and W are real Euclidean spaces.

(Refer Slide Time: 21:04)



So let us consider the case when V is R n and W is R m in which case this will reduce to a matrix equation this equation will be replaced by the matrix equation A x equal to b I want to solve the equation A x equal to b you can think of A as being the matrix of T relative so some basis think of A as being the matrix of T relative to some given basis for R n and R m then you know that this is consistent if and only if this is consistent, range space of T is range space of A so I can think of A as matrix as well as a linear transformation ok, this is the usual practice in linear algebra.

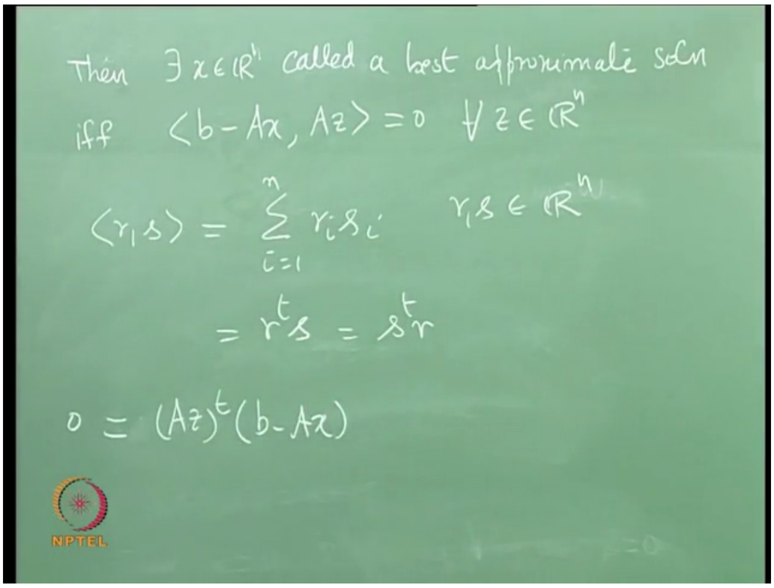
The question is does this system have a solution? I mention that there are situations where many situations where this does not have a solution in which case we would like to minimize the norm of A x minus b, so we are looking at the following problem, does there exists, the problem is norm A x minus b, I want this I want an x in R n such that norm A x minus b is less and or equal to norm let us say A Z minus b, x must be fixed so this is fixed this is like U and this is like U this is like x, so lets try to reduce this problem to the problem of best approximation, b plays the role of x in the previous instance A x plays a role of U I want to find an x such that norm A x minus b is less and or equal to norm A Z minus b for all Z and R n.

See A is a linear transformation from V in W so A is an m cross n matrix, the domain of A is R n codomain in R m so A is an m cross n matrix, so Z must be from R n Z is a variable x is fixed b is fixed, b is like x, A x is like U, this is like W, norm U minus x less and or equal to norm U minus W sorry, norm U minus x less and or equal to norm x minus W ok, instead of x

I have b, instead of W I have A Z small w comes from capital W that is a subspace, instead of capital W I have range of A is that clear? So let me write then, this is like U, this is like x, this is like a general w from capital W and so I will write capital W as range of A, this b is offcourse is x, b is a fixed vector. I want to minimize norm b is a fixed vector like x, so I want to minimize U minus x.

So among all vectors in the range of A, I pick that particular vector U which is equal to A x so the question balls to balls down to finding x, I need to find an x in R n that satisfies this inequality these inequality.

(Refer Slide Time: 24:43)



The previous theorem says that this problem has a solution if and only if certain condition is satisfied that is x minus U must be orthogonal to W then there exist x in R n that satisfies ok, let me give a name to this called a best approximate solution, it is called a best approximate solution best approximation from the previous discussion this time x is called a best approximate solution.

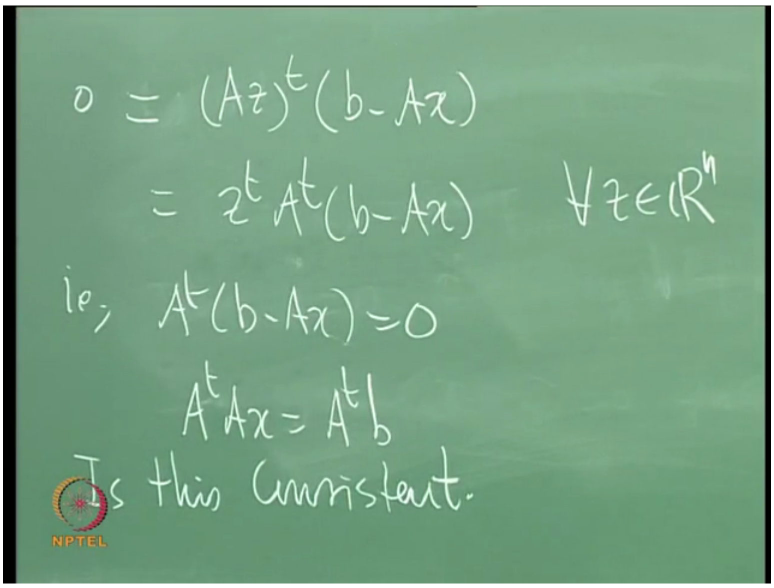
It minimizes norm A x minus b, if and only if x minus U perpendicular to W in the previous instance, instead of x I have b, instead of U I have A x that is in the range of A and W belongs to capital W capital W I have written that is the range of A. So this b minus A x must be perpendicular to let me write A Z, this must be zero for all Z and R n this is the condition coming from the previous theorem, x is a best approximate solution that is among all vectors A Z replace Z by x then norm A x minus b is the least among all the numbers norm A Z minus

b. Look at this condition once again, I am in the situation when the space is R n so this is the usual inner product and so I can write this in terms of the formula.

What is the inner product between two vectors let us say W, Z ok let us say r, s is summation ri si if r and s are in R n. I can write this as r transpose s where I always write when I write a vector standing alone it will be a column vector it will be a row vector, any vector I write in the situation will be a row vector so r is row vector r transpose is, is that correct? I go back any vector standing alone will be a column vector, so this r and s are column vectors, r transpose is row vector, row into column that is precisely this, ok.

So this inner product can replace by this formula I will use that here, zero equals b minus A x by the way r transpose s is also equal to s transpose r in the real case ok this is what I want really, s transpose r, transpose of this times this, A Z transpose times b minus A x is this clear? I am replacing the see this is a usual inner product I am replacing the usual inner product by the formula involving transposes then this reduces to this you expand this transpose satisfies, see we have seen this operation of transpose earlier this satisfies the reverse order law.

(Refer Slide Time: 28:09)



So this is Z transpose A transpose b minus A x this must hold for all Z in R n which means you can again think of the inner product.

The inner product of Z with this vector is zero for all Z the inner (pro) see this is again a kind of a transpose Z transpose into some vector see this is a vector in R n ok please check this is a vector in R n, A transpose of something see A of something is in R m A transpose of something is in R n so this a vector in R n ok. So compatibility there is no problem Z is in so you look at this vector this is a vector I am sorry this is a vector that is perpendicular to all vectors, so the vector must be zero that is A transpose into b minus A x must be zero that is if x is a best approximate solution ok that is x is a best approximate solution if and only if x satisfies this equation ok, this intermediate variable U has been removed that is Z has been removed.

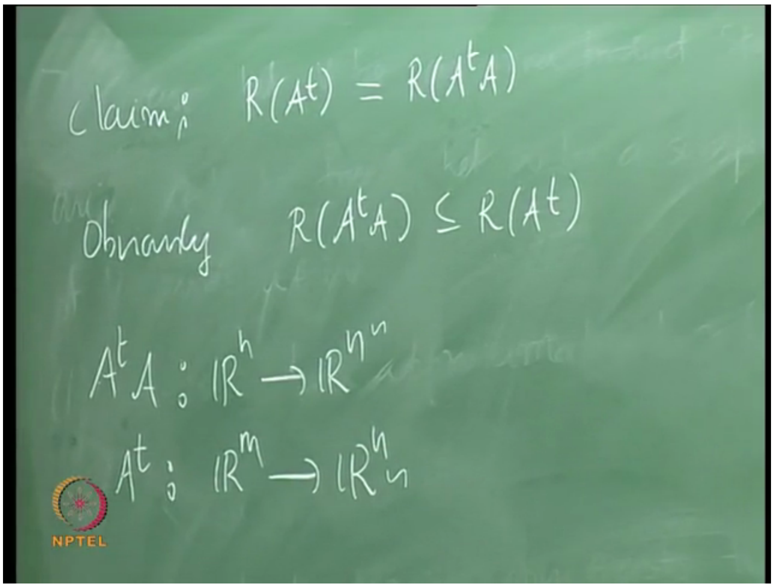
This is a necessary sufficient condition coming from the previous theorem for x to be a best approximate solution of the system A x equal to b, if A x equal to b, if the system is consistent this is obviously satisfied ok, but if its not satisfied then this condition must hold for that x ok if x if A x minus b is not zero then A x minus b must satisfies this, but what is this equation? You rewrite this as A transpose A x equals A transpose b, does this system have a solution? That is the problem, x is a best approximate solution if and only if x satisfies this equation but is this system consistent?

We need to consider that question also, is this system consistent? Does this system have a solution? See given b I can calculate the right hand side I can calculate A transpose A, A is known the question is, is this consistent? Is this system consistent? Ok do you have an answer? Can you rewrite reformulate this question? See what did I say with regard to this equation, lets say the same question can be asked here, when is the system consistent? If b belongs to range of A ok, the question is, look at the vector A transpose b, does it belong to range of A transpose A? Does the vector A transpose b belong to the range of A transpose A, that is the question. What is the answer?

See the answer is not really straightforward so I don't expect you to give the answer, the answer is yes ok the answer is yes and look at the full picture see let me go back the answer is yes and look at the full picture, what is the full picture? A x equal to b in general does not have a solution, there is no x that satisfies this equation, but there is always an x that satisfies this equation ok. So from an inconsistent system we have come to a consistent system, you remember that we have gone from one system to another we cannot go back. In general we cannot go back so there is no contradiction here.

See from this we have come to this, but you can't in general go back and so this two systems are not the same ok, this two systems are not the same if the systems are the same then there is a contradiction, this is inconsistent whereas this is consistent ok. Just think it over but I want to show this is consistent always, I want to show that this is system is consistent irrespective to of what b is. So the proof come from this equation connecting subspaces that takes a little effort to prove.

(Refer Slide Time: 33:09)



So I want to make the following claim, look at range of A transpose I want to show that this is equal to range of A transpose A, range of A transpose is equal to range of A transpose A.

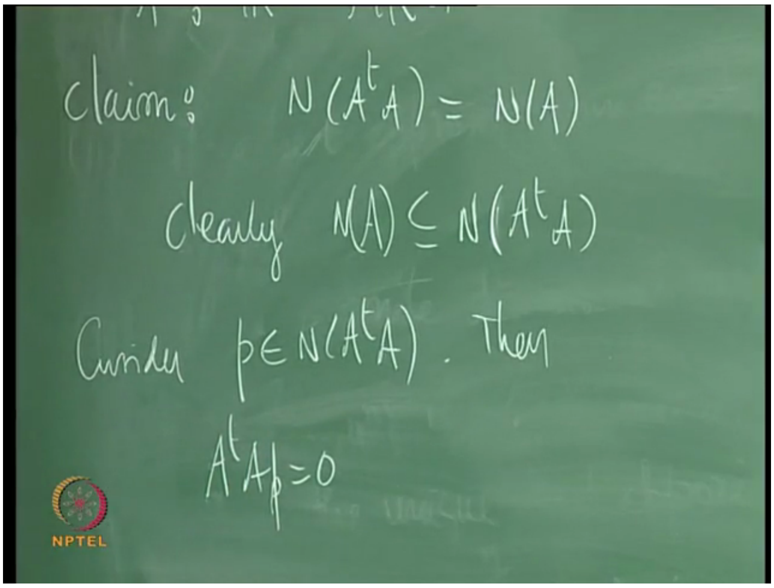
We have not seen this equation before, one of the inclusions is straightforward, range of A transpose A contained in range of A transpose that is clear. It is the other way which requires a little effort. Suppose I prove this ok, then you can go back to this equation see that this is consistent right hand side vector belongs to range of A transpose the coefficients matrix is here is A transpose A. So A transpose b belongs to range of A transpose A and so this systems is consistent ok. So will prove this equation and then it follows that system is consistent which means that there is always a best approximate solution for linear equations ok.

Given a n equation A x equal to b if it is consistent there is no problem if it is not consistent we can always find the best approximate solution where the best approximation is respect to the usual norm that is when I want to minimize norm A x minus b I am looking at the Euclidean norm ok. So I need to prove this, I told you that one of the inclusion is obvious. So let me record that, obviously range of A transpose A is contained in range of A transpose, by the way you also need to verify that this equation makes sense that is we are talking about two subspaces in the same vector space, is that first clear?

See ok lets make a quick check of that, A transpose A is from R n to R n, A transpose A is a square matrix if you look at A transpose it is from R m to R n, A is from R n to R m, A transpose is from R m to R n so if you look at range of A transpose A that is here in R n range of A transpose is in R n. So first of all this equation is you know is you can hope for the validity it is well defined to talk about that ok, this is true, what I will show is that I must show the range of A transpose is contain range of A transpose A instead I will show that the dimensions go inside, so I have two subspaces one contained in the other dimensions being the same the subspaces must be the same ok.

In order to prove that the dimension is the same I will use Rank Nullity dimension theorem but before that I need the following.

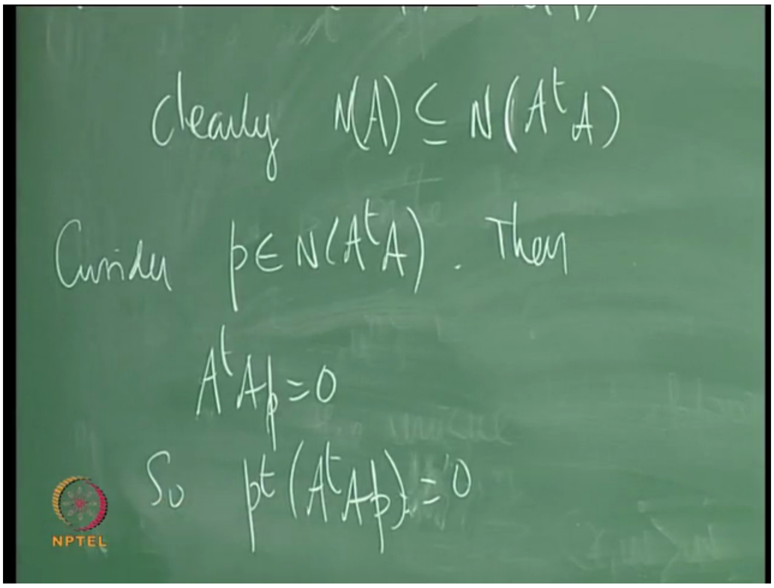
(Refer Slide Time: 36:00)



To prove this claim I will prove another claim which is null space of A transpose A equals so can you make a guess of what this is, range of A transpose A is equal to range of A transpose is null space of A only then you have compatibility again you can check that this makes sense. A transpose A is null space of that in R n A is from R n to R m so null space of A is contained in R m both this are subspaces of R m, this is easy to prove. Again one of them is clear, which one? If A x equal to zero then A transpose A x equal to zero so null space of A is contained in null space of A transpose A ok one inclusion is straightforward.

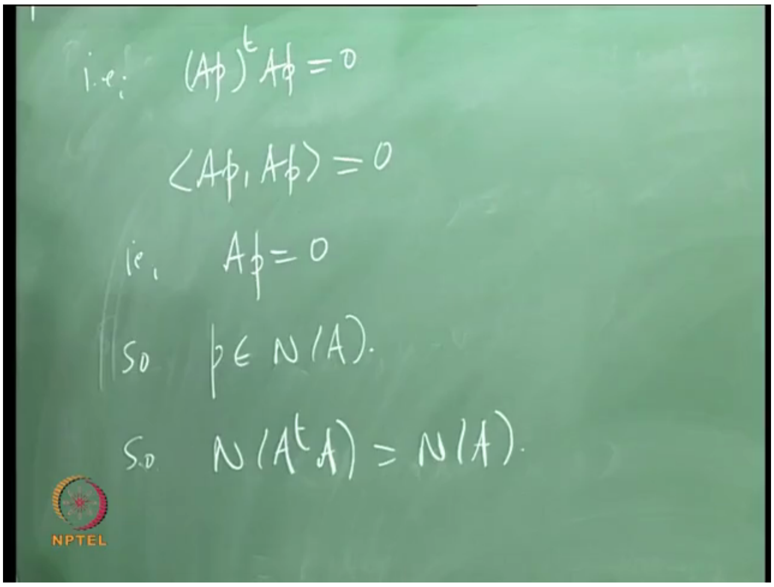
That is if A x equal to zero then A transpose A x is zero so if x belong to null spaces of A then x belongs to null space of A transpose A. I need to prove the other way around. So consider lets take some vector different from x say P, suppose P belongs to null space of A transpose A then A transpose A P is zero, this is a zero vector, zero vector is perpendicular to any vector in particular to P.

(Refer Slide Time: 37:26)



So P transpose into A transpose A P, this must be zero. P transpose into A transpose A P is zero.

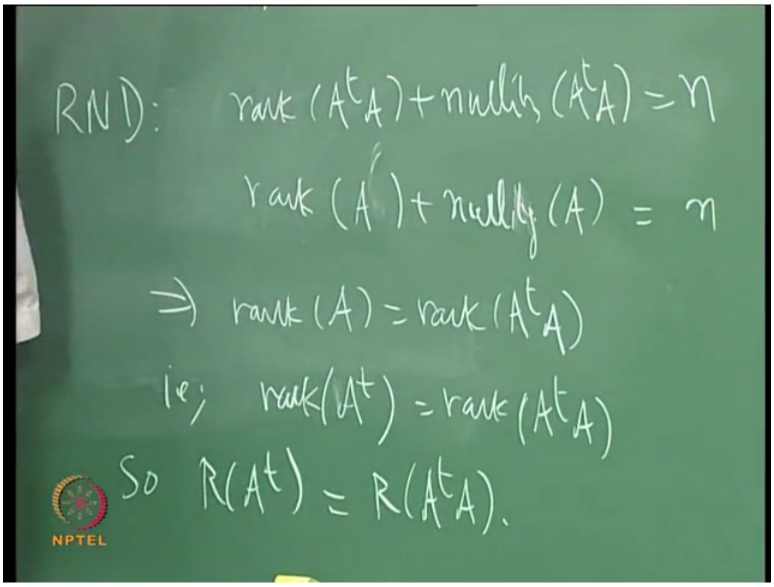
(Refer Slide Time: 37:45)



But this can be rewritten as A P transpose A P but that is the same as saying inner product of A p with itself is zero where R n using the standard inner product that is A P is zero.

So we started with A transpose A P is zero we have proved A P is zero so it follows that P belongs to null space of A, so null space of A transpose A is equal to null space of A, apply Rank Nullity dimension theorem and also apply the fact that the rank of a (mat) the row rank and the column rank are the same ok. So lets apply Rank Nullity dimension theorem for A transpose A first.

(Refer Slide Time: 38:40)



Rank Nullity dimension theorem for A transpose A, rank of A transpose A plus I will write it in full nullity A transpose A must be the dimension of the domain space that is n. For A transpose rank of A transpose ok rank of A plus nullity of A equal dimension of the domain space, I am applying it to A.

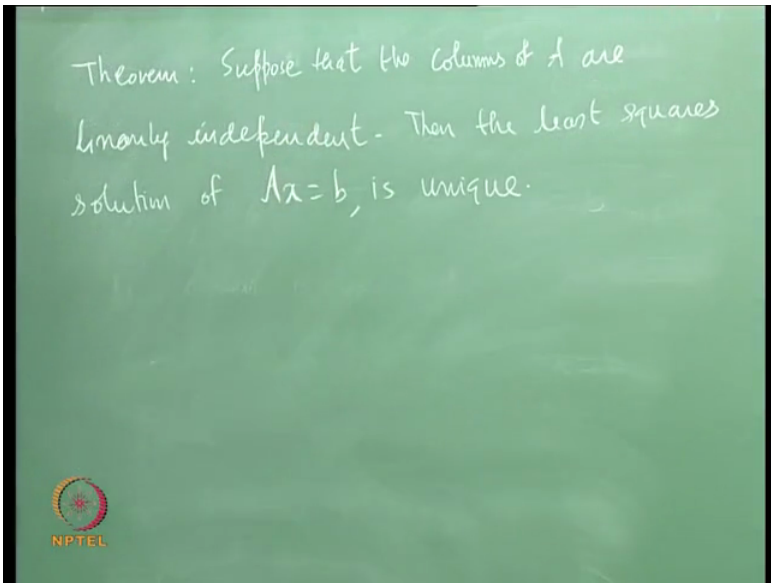
Rank of A plus nullity of A is a dimension of the domain space, what I have shown is that null space of A and null space of A transpose are the same. So I subtract one from the other this two get cancelled, it means rank of A is equal to rank of A transpose A, you can't apply range immediately because range of A is in R m ok range of A is in R m but rank of A is rank of A transpose so that is what you need to apply then conclude that is, rank of A is rank of A transpose this is the same as saying that the row rank is the same as the column rank.

So rank of A transpose is equal to rank of A transpose A now apply the fact that the dimension, what is the rank of A transpose A? It is the dimension of the range of A transpose A, go back to this equation. So range of A transpose A range of A transpose they are the same they are subspaces of the same vectors space the ranks coincide and one is contained in the other. So they must be equal. So range of A transpose is equal to range of A transpose A, ok so it is with this little extra effort using the Rank Nullity dimension theorem one could show that the system A transpose A x equal to A transpose b is consistent.

You might have come across this equation in numerical analysis, this is what is called as normal equation this equation are called normal equations in the problem of for instance interpolation ok so in the normal equation are always consistent, so that is the complete answer then. This gives us a method in principle if you want a best approximate solution of this equation by the way we have to go back to the familiar least square solution ok because our norms are the Euclidean norms some sub-squares or integrals of squares, so it is a least square solution. If the system A x equal to b is inconsistent then one seeks least square solutions.

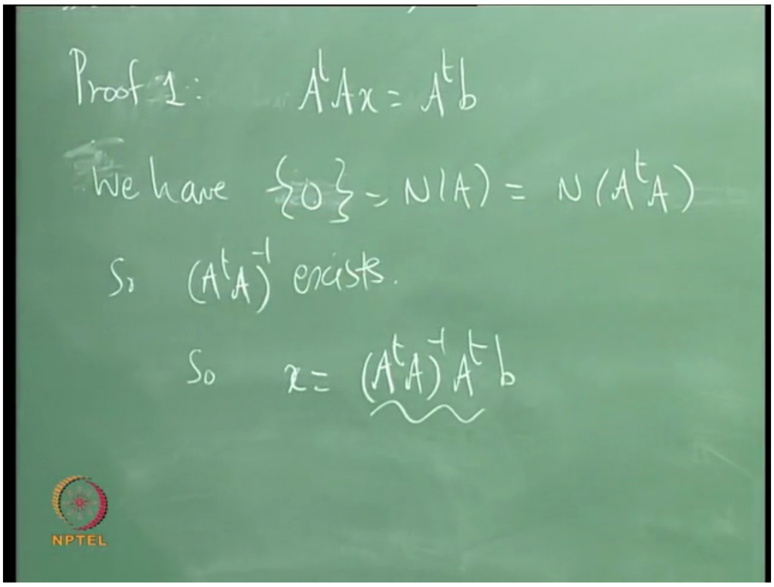
Least Squares Solutions always exists, that is the same as saying this system is consistent ok, least square solutions always exists the questions whether the least square solutions are unique, let me just give you an instance where the least square solutions is unique and stop.

(Refer Slide Time: 42:25)



So I have the following result, suppose that the columns of A are linearly independent, then the least squares solution of the system A x equal to b is unique. I will give two proofs, one uses a principle the other one will be useful in practice.

(Refer Slide Time: 43:27)



First proof, proof 1, I am looking at the equation A transpose A x equals A transpose b when the column of A are linearly independent what is the null space of A?

And the columns of A are linearly independent, what is the null space of A? Single term zero, null space of (we) have shown to be equal to something else some other subspace, what is that? Just now we have proved, null space of A transpose A the difference between A and A transpose A is if even if A is (rectang) whatever be A , A transpose A is square, A transpose A is a square matrix whose null space consists of single term zero, so A transpose A as a linear transformation is injective but since its square it must be surjective so it must be invertible So A transpose A inverse exists, just pre-multiply by A transpose A inverse to get x to be A transpose A inverse A transpose b.

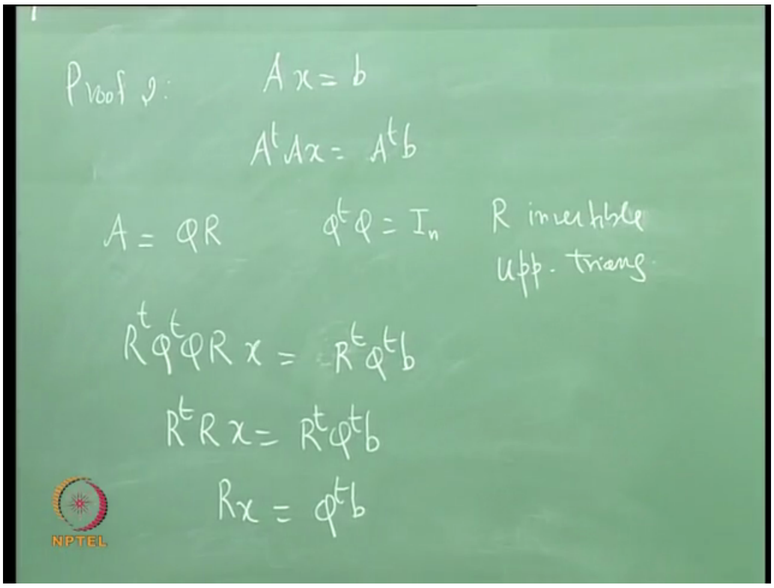
This is unique because this is like this is a system where the coefficient matrix is invertible so the system must have unique solution ok. That unique solution is given by this formula, we will come back and look at this formula A transpose A inverse A transpose. We will discuss a notion of generalized inverse in place of this ok that will be done a little later. This is a proof which uses a principle that null space of A transpose A equal to null space of A. In practice, one could use the Q R decomposition. So my second proof is by using the Q R decomposition ok, before I give the second proof let this be clear.

A x equal to b if it is consistent the question of finding a least square solution does not arrives, so this question arises only if A x equal to b is inconsistent if A x equal to b is inconsistent and if the columns are independent there is a unique least square solution, if not there are infinitely many least square solutions. In general there are infinitely many least square solutions. In such a case which solution would you be satisfied with? Infinitely many least squares solutions you should be satisfied with one solution which for which a suggestion is to locate a vector which has the least norm among all those vectors ok this problem will discuss later.

For instance if you have a even in the case of a rectangular system of consistent equations there are infinitely many solutions. Rectangular system consistent infinitely many solutions you should be satisfied with one with a least norm the same role applies for the inconsistent case also, A x equal to b inconsistent but it has least square solution. In general there are many infinitely many least squares solution there is a situation when the solution is unique, so if you go back to the problem when there are infinitely many least square solutions you should be interested in the minimum norm least norm of all those solutions.

The question of whether such a vector exists we will discuss that later ok, just to give a little more enlarged picture I wanted to make this comment so I will just go back to the second proof and tell you how the Q R decomposition could be used.

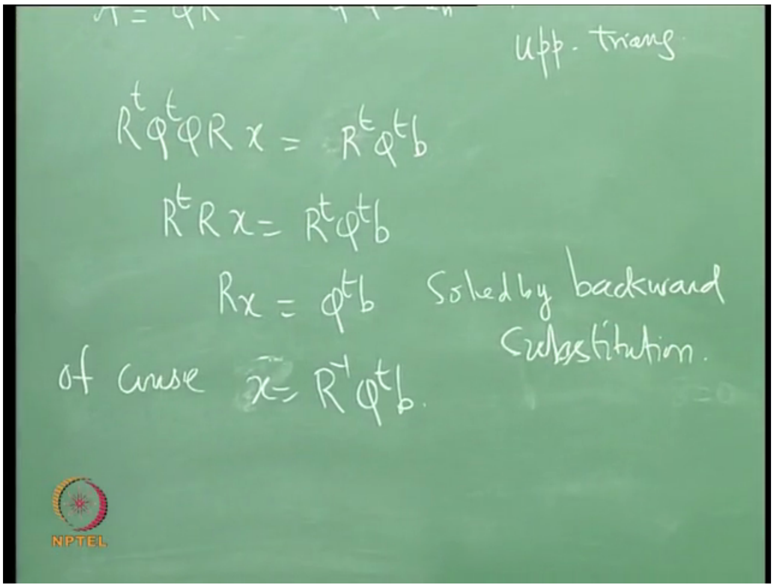
(Refer Slide Time: 47:34)



Second proof, I have A x equal to b and this from this I get A transpose A x equals A transpose b ok that is the equation corresponding to the normal equations, A is Q R this is applicable because the columns of A are independent ok remember the framework in which we derived the Q R decomposition the columns of A must be linearly independent that is the situation here.

So this is possible A equal to Q R with Q transpose Q being identity of order n R is invertible and upper triangular so back and substitute here, A transpose A is ok I will write down this quickly R transpose Q transpose Q R that is the left hand side equals R transpose Q transpose b, Q transpose Q is identity R transpose R x equals R transpose Q transpose b. R square invertible so at R square invertible R transpose is also invertible so this is the same as R x equals Q transpose b.

(Refer Slide Time: 49:09)



From this one can write x is equal to R inverse Q transpose b ok. Offcourse that is possible but then one would like to stop with this and then look at the structure of R and see what you must do to solve this system. R is upper triangular so the last equation has only one variable the last but one equation has two variables etc. So we do what is called as a backward substitution, find the variable x n from the last equation go to the previous one find x n minus 1 etc. The first equation gives x1 ok, so this is solved by backward substitution ok, so this is something that could be applied in a numerical example whereas this gives a conceptual proof ok, so let me stop.