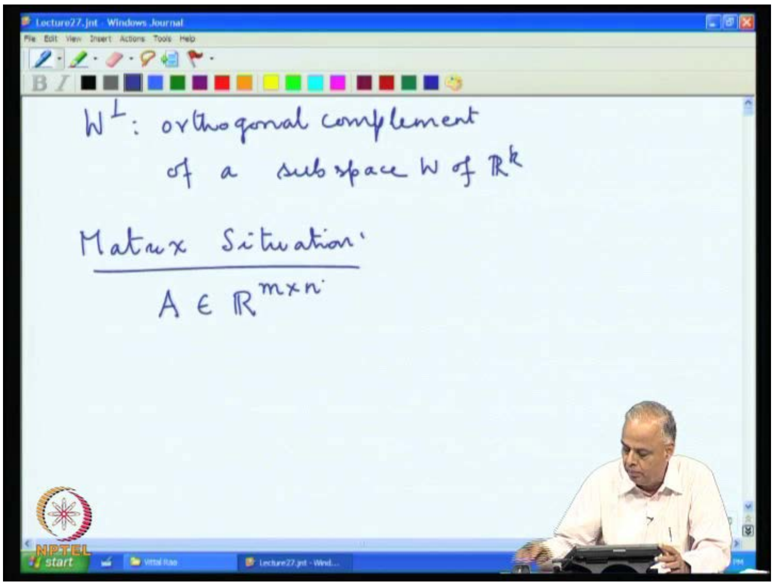
Advanced Matrix Theory and Linear Algebra for Engineers
Prof. R. Vittal Rao
Centre for Electronics Design and Technology
Indian Institute of Science, Bangalore

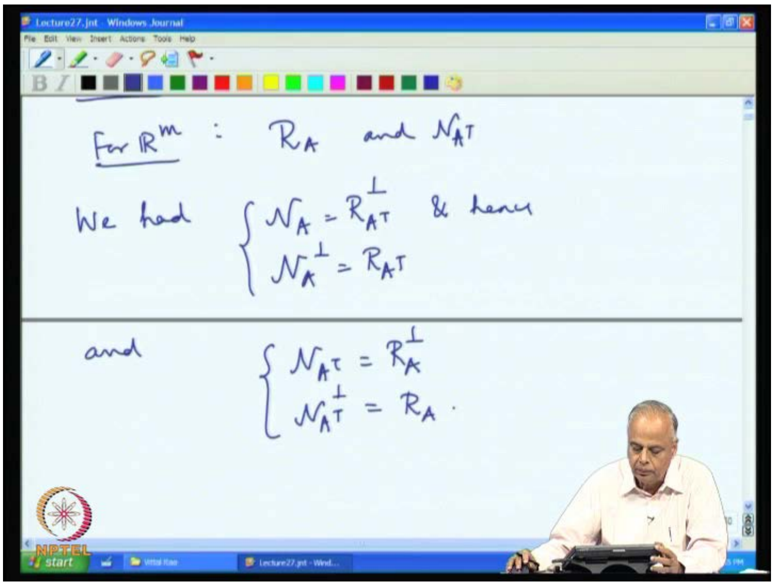
Lecture No. # 27
Inner Product and Orthogonality- Part 6

(Refer Slide Time: 00:23)



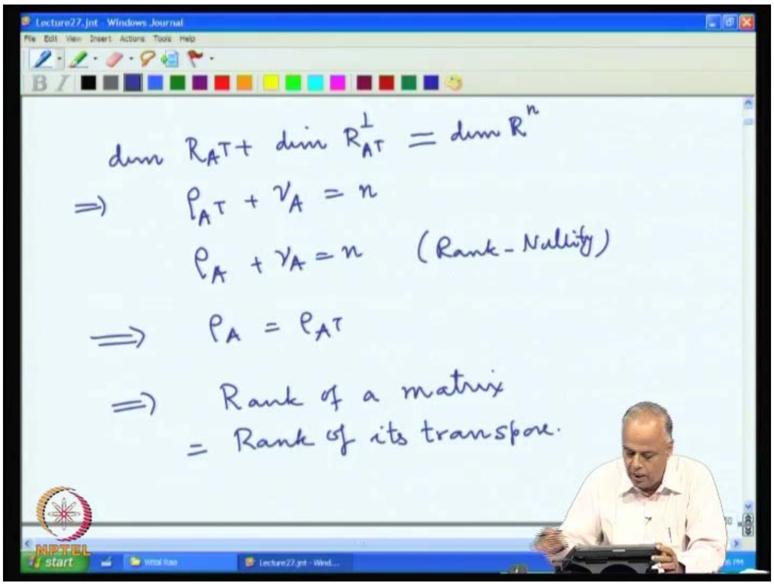
We have looked at, the notion of the orthogonal complement of a subspace in R k, orthogonal complement of a sub space W of R k. And then, we use the properties of this orthogonal complement, in the case of the matrix situation. In the matrix situation, we have a matrix A, which is an m by n matrix real and we have on the for R n.

(Refer Slide Time: 01:00)



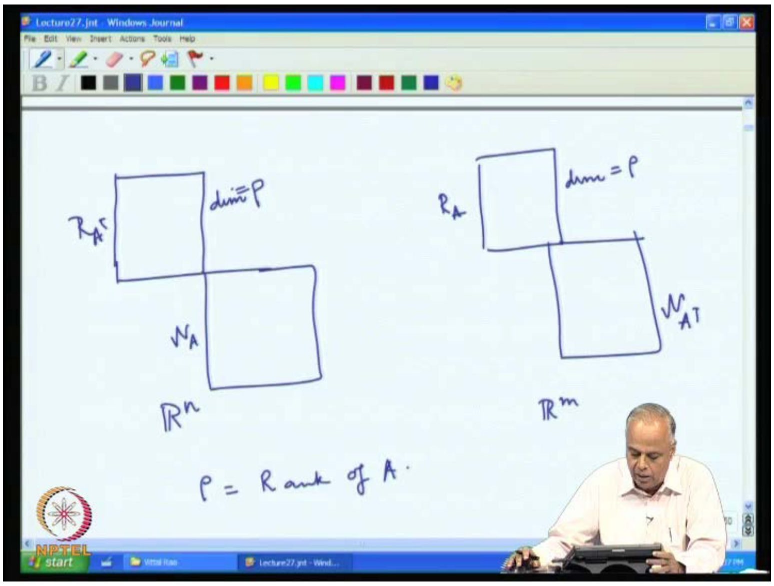
We have 2 subspaces namely, the range of A transposes and the null space of A, and for R m, we have the two sub spaces, the range of A and the null space of A transpose. We observe, that these 2 pairs of subspaces or orthogonal to each other. So, we had the null space of A, if the R A transposes perpendicular and hence, the null space of A perp is R A transpose and similarly, on the R m side, the null space of A transpose, if the range of A perpendicular, and the null space of A transpose perpendicular is R A. We then, used the notion that, the dimension of any subspace plus is orthogonal complement is the whole space.

(Refer Slide Time: 02:06)



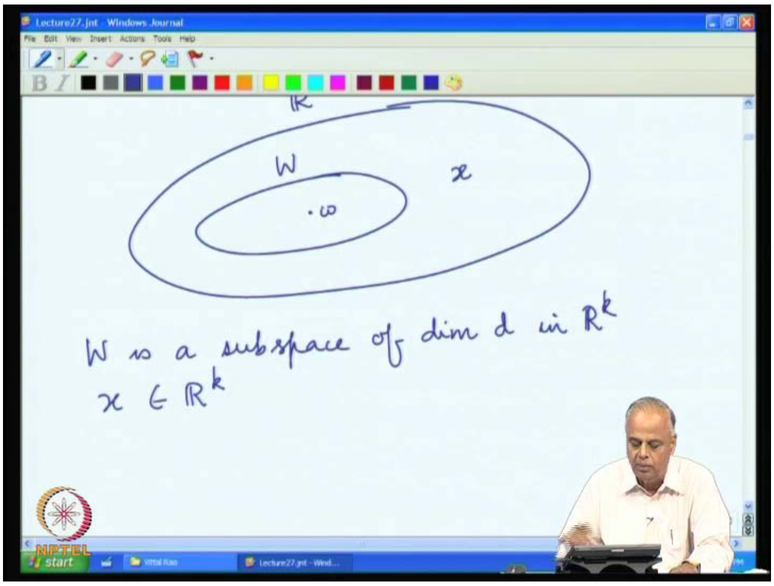
So, dimension of R A plus dimension of R A transpose perp So, the dimension of R A transpose plus the dimension of R A transpose perp. Since, both are subspaces of R n must be equal to the dimension of R n. Which gave as, this is the rank of A transpose and since, R A transpose perpendicular is equal to N A. This dimension of same as dimension N A, which is the nullity and that was equal to n. And by Rank Nullity theorem by rank the nullity theorem we had rho A plus nu A equal to n. These two together gave us that, rho A equal to rho A transpose. Therefore, the rank of a matrix is equal to it is transpose, so the rank a rank of a matrix is equal to rank of it is transpose. This is the first important thing that we observe.

(Refer Slide Time: 03:31)



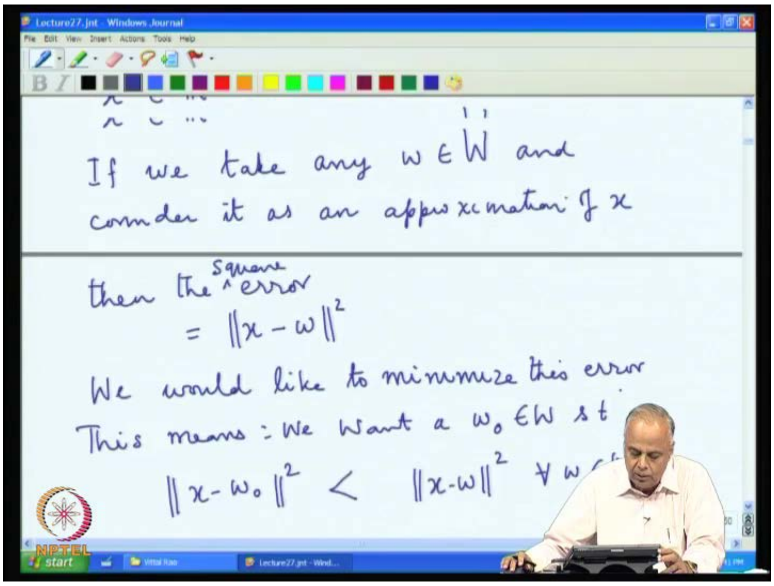
Therefore, we had on the R n side, the two orthogonal subspaces each being complement orthogonal complement of the other. On the R m side, we had this null space of a transpose and the range of A and again, these 2 pairs of subspaces are orthogonal to each other and the dimension of this and the dimension of this are the same, because the rank of A is rank of A transpose found and now, we denote by the common dimension s rho. So, rho is the rank of A. Now, want we will not like rows of A, because rows of A is same as rows of A transpose. The dimension here is nu A and the dimension here is nu a transpose. As we mentioned in the last lecture, have a basic aim will be to find orthonormal basis to the 4 subspaces in such a manner, that it helps our problem of analysing the matrix.

(Refer Slide Time: 04:52)



We shall now, look at 1 more property of the sub space, that is what is known as the best of approximation from a subspace. The best approximation from a subspace So, what do we mean by best approximation? We have the space R k and in that, we have a subspace W and let us say, W is a subspace of dimension d in R k. So, we have a subspace of dimension d in R k and now, what we have going to do is, we look at a vector x which is outside W, it may be inside W, but we take a general vector x in R k. Now, what we would like to do is try to approximate this vector x with a vector W in W.

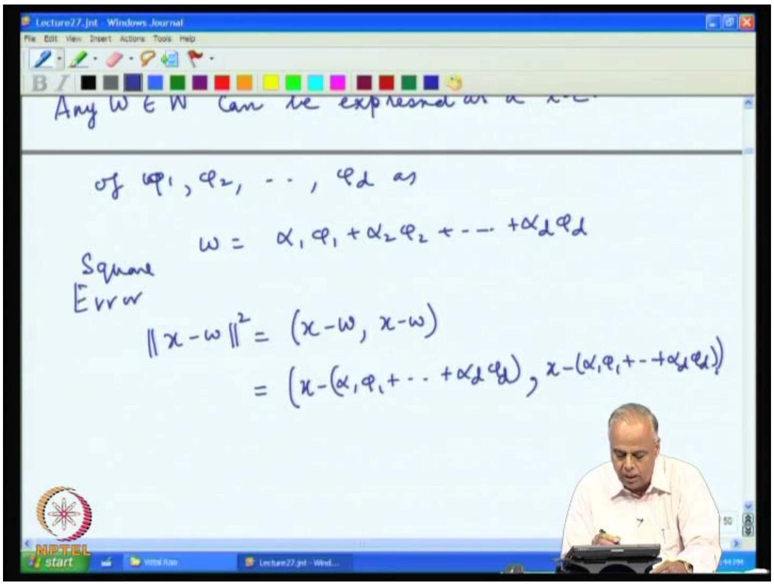
(Refer Slide Time: 06:04)



So, if we take if we take any w in W and consider it as an approximation of x. Then the error is quantified, we wanted x be a approximately the w, so there is going to be an error and that error is going to be quantified by this square of the length of the difference vector. So, we wanted x, we had approximating it w. So, the error vector is x minus w, this is called the square error. this is called the square error The square error is the length of the error vector squared. Then, the error we get is this, we would like to minimize this error so, we would like to minimize this error. What do we mean by this? This means: we want a w 0 in W such that, if you now take w 0, as the approximant and calculate the error.

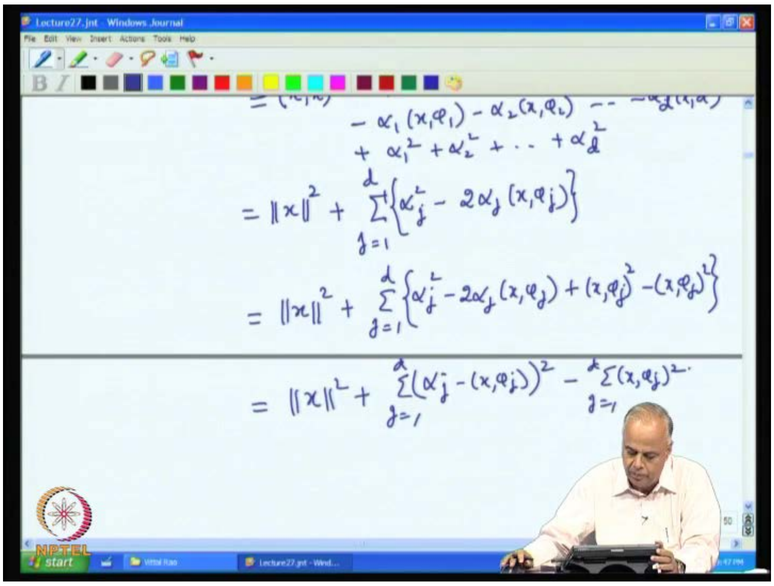
This is the error, that is got by taking w 0 as the approximation of x. Now, we look at the error, I get by taking any other w 0, that it be x minus w square. This must be, strictly less than this for every w in W; W not equal to w 0. This means if, you take the vector w 0 and calculate the error, the error is going to be smaller than, by taking any other vector at the approximant for x. And hence w 0 becomes the best approximation with the mean square, with the square of the error being least. So, thus call the least square approximation. Does there exist such a w 0? Obviously, the way we have stated, it means the w 0 unique, because the moment w 0 every other w is going to get more error. So, we are expecting that there will be a single vector w 0, which will approximate. So, we would like to know, whether there exist such a w 0. Let us investigate; this question now, W is a dimension d and therefore, we can find an orthonormal basis and how many vectors will it contain? It will contain exactly d vectors.

(Refer Slide Time: 09:15)



So let phi 1, phi 2, phi d be an orthonormal basis for W. Now we have the d dimensional space w and we have found an orthonoramal basis. Now, if you take w any vector in W. So, any vector w in W can be express as a linear combination of phi 1, phi 2, phi d because phi 1, phi 2, phi d is a basis for w, can be expressed as a linear combination of phi 1, phi 2, phi d as, w equal to say alpha 1, phi 1 plus alpha 2 phi 2 plus alpha d phi d. And therefore, the error will be square error, x minus w squared, the length of the vector square is the dot product of x minus w with itself the length of any vector square is the dot product of the vector with itself or the inner product or whichever we want to call it. So, which is the same as, the dot product of the vector alpha 1, alpha d phi d with the same vector. So, this is the inner product.

(Refer Slide Time: 11:00)

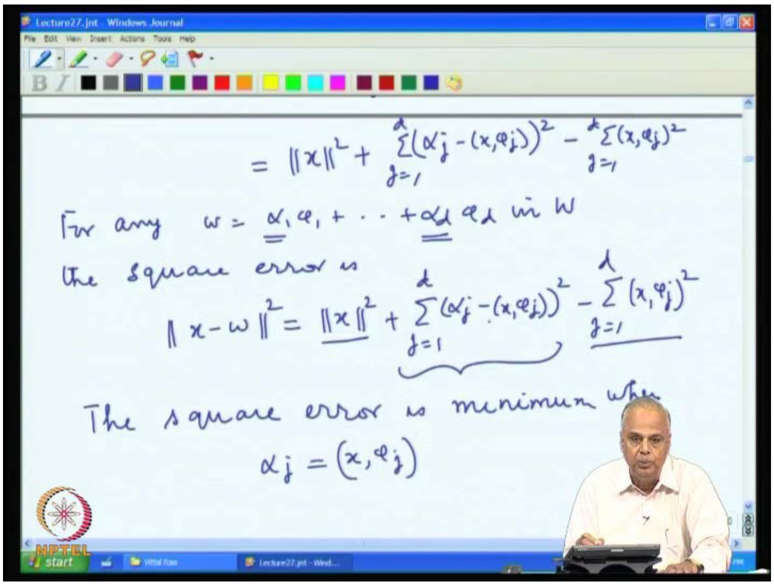


Now, we know that the dot product is distributed we can take term by term inner product, so the first 1 gives me x comma x, the second 1 gives me minus alpha 1 x comma phi 1 minus alpha 2 x comma phi 2 and so on minus alpha k x comma phi k. Then, we get another from here minus alpha 1 taking these term again we get minus alpha 1 x comma phi 1, we get phi 1 comma x 1. But since, we are in the real space the dot product is commutative x dot u is same as u dot x.

So, again we can write it as x comma phi 1 x comma phi 2 and x comma phi d. Then finally, we get the cross products of alphas since phi are orthonormal, the phi i phi j terms will go and the phi i phi i terms will give 1, so i will get plus alpha 1 squared plus alpha 2 squared plus alpha d squared. We will write this, in summation notation x comma x is first of all the length of squared and this remaining thing can be written as, summation j equal to 1 to d alpha j squared, that this (( )) squared terms minus 2 alpha j x phi j that take square of the remaining terms. So, we have this.

Now, therefore the square of the error, if I take any vector w in W and calculate this square of the error and this square of the error, turns out to be this. You make a simplification; we complete the square inside the sum, the alpha j squared minus 2 alpha j x phi j, which is of the form a squared minus 2 a b, so if we add b squared and subtract this square in order to complete this square. This is now, equal to plus summation j equal to 1 to d alpha j minus x phi j the quantity squared minus that extra thing, which we added in subtracted x phi j squared.

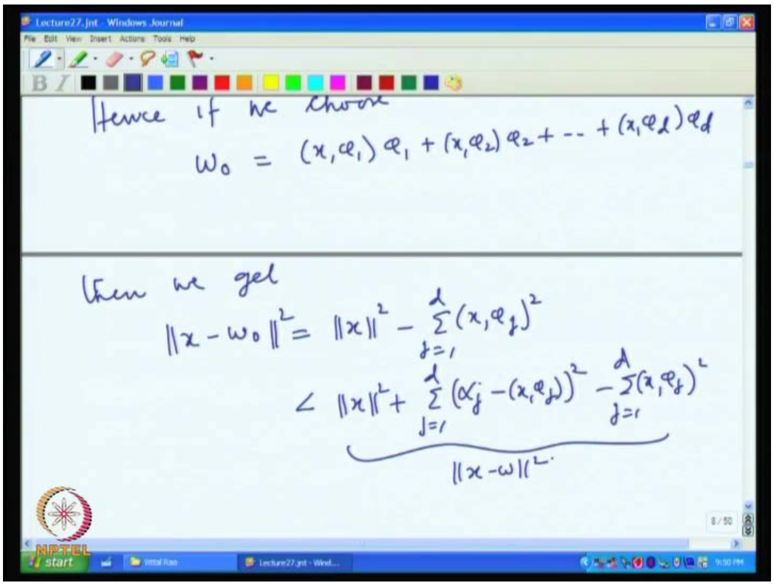
(Refer Slide Time: 13:42)



Now therefore, for any w equal to alpha 1 phi 1plus alpha d phi d in W. The square error is norm x minus w squared, the length squared is equal to the length of x squared plus j equal to 1 to d alpha j minus x phi j squared minus sum of j equal to 1 to infinity x phi j squared. Now, we are trying to minimize the error when, we are saying that we are minimizing the error. This means, we are minimizing the right hand side we are minimizing the right hand side by the vector x is given, we can do nothing about it and the orthonormal set in w space.

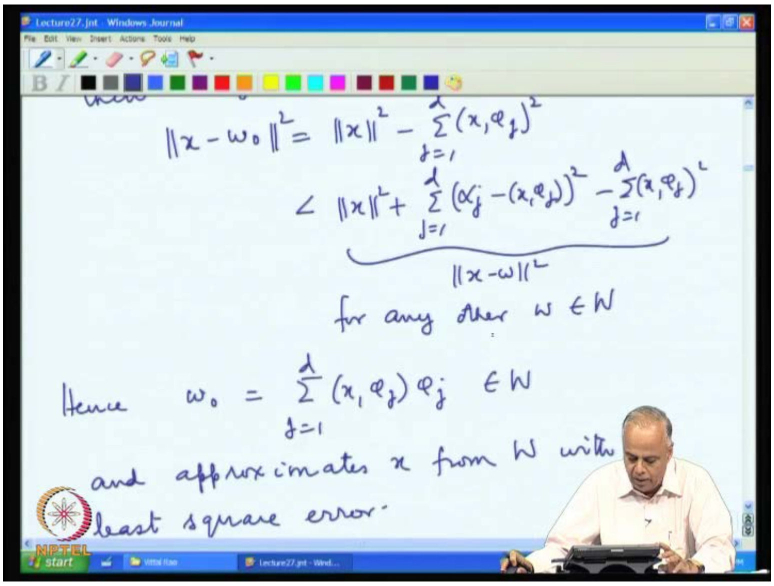
So, we can do nothing about it. We are trying to minimizing by choosing the suitable vector in w, by choosing a suitable vector in w. We mean suitable values for alpha 1 alpha 2 and alpha d. This means, we have to adjust the alphas in such a way the right hand side becomes minimum the alphas appear in the middle term. Now, if you have to make it minimum. This is the sum of squares. So, each term is nonnegative, so the best way to make it minimal it by making each term zero. So the error the square error is minimum the square error is minimum when alpha j is equal to x phi j. If any one of the alpha j is not equal to x phi j then, there will be an additional contribution from the middle term and therefore, the error will increase.

(Refer Slide Time: 15:49)



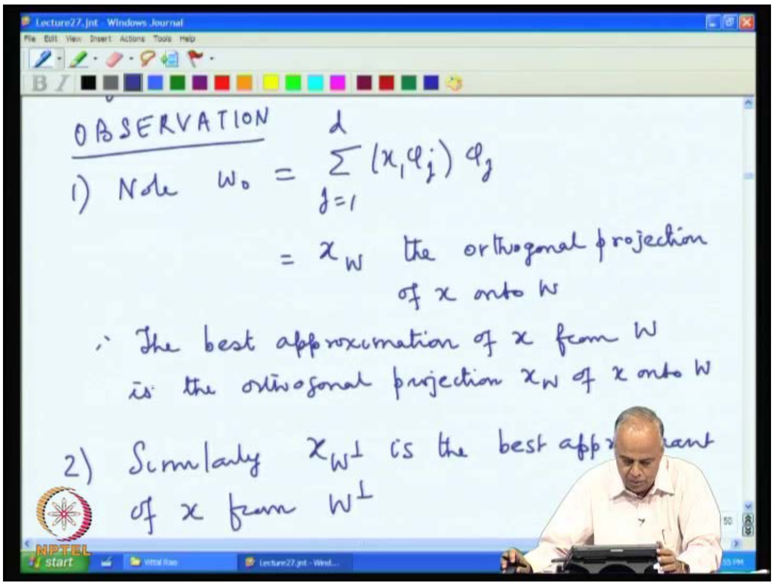
Hence, if we choose w 0 to be equal to by choosing these values of alpha 1 alpha 2 alpha d, we get x phi 1 phi 1 plus x phi 2 phi 2 plus x phi d phi d. Then, we get the error is obtained exactly as norm x squared minus summation j equal to 1 to d x phi j squared. This is strictly, less than norm x squared plus summation j equal to 1 to d alpha j minus x phi j squared minus summation j equal to 1 to d x phi j squared, for which is the same as this quantity is norm x minus w squared for any other w in W.

(Refer Slide Time: 17:14)



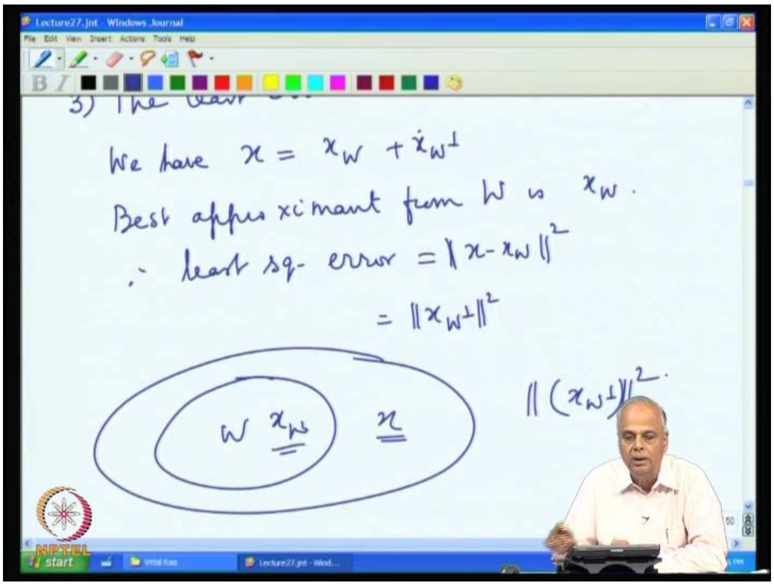
And therefore, w naught with the vector that is the least here. hence w 0 equal to summation j equal to 1 to d x phi j phi j belongs to w and approximates x from W with least f with least square error. This is call the best approximation this is called the best approximation of x from W. We are going to make a few observations now, so what we have what so far is if you start with a vector space, if the vector space R k and then you look at, this is the vector space R k inside that vector space if the subspace w and we are looking for any given vector x in R k, a vector inside w which approximates it with the least error least square error. And we are now found that least square error approximation is given by this vector w naught.

(Refer Slide Time: 18:47)



Now we are going to make certain observations. You recall that the right hand side of w naught the the vector w naught has been defined to be this and if we recall this is precisely the orthogonal projection of x on to w. So note, w 0 is equal to j equal to 1 to d x phi j phi j is equal to x w, the orthogonal projection the orthogonal projection of x on to W. Therefore the orthogonal projection is the best approximate. So therefore, the best approximates the best approximation of x from W is the orthogonal projection x w of x on to w. Similarly, the orthogonal projection of x on to w perp is the best approximate of x from w perp. Similarly, x w perp is the best approximate of x from w perp. Notice, that if you know that x w we also know x w perp because x w plus x w perp is x and therefore, if you know 1 of the 2 approximants, we automatically know the other approximants. Therefore, we know how to get this best approximation, what it the least error? The least error that we get is given by the error that we get by taking w naught as the approximation which is norm x squared minus j equal to 1 to d x squared d squared.

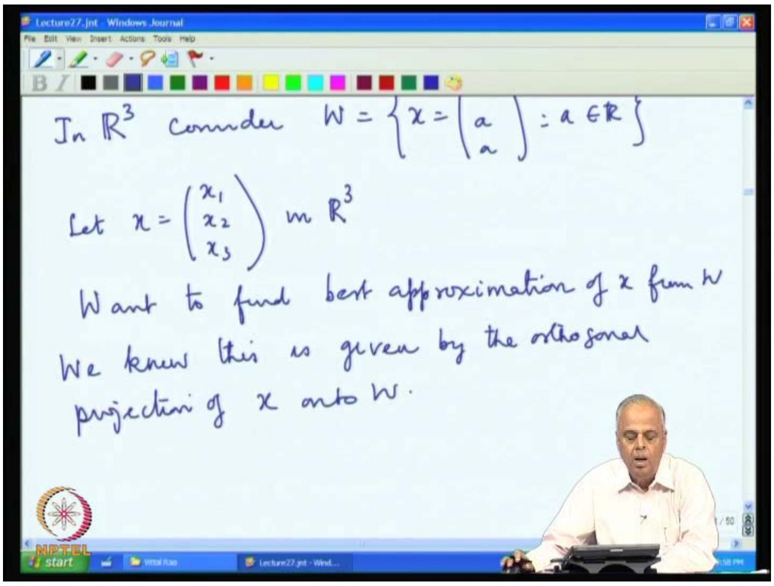
(Refer Slide Time: 21:20)



So the least error is least square error is equal to summation j equal to 1 to d x phi j squared, if you now look at the error this is w naught x minus let me let us do this derivate more carefully , let us first find what is we have x is equal to x w plus x w perp and therefore, the best approximant we have seen from w is x w therefore, least square error is equal to x minus x w squared, but Pythagoras theorem told us that x w squared plus x w perp squared in norm is equal to the norm of x squared, so this is equal to norm x w perp squared. Since, because x minus w is equal to x w perp.

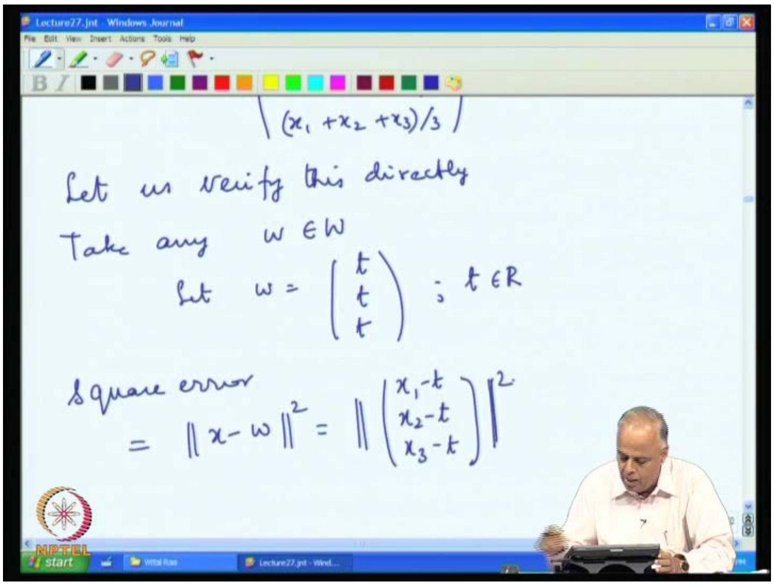
So, we have the least the error is the projection on to the other orthogonal complement. So, if you take w if you take w and if you take x and if you look at x w and this will be the best approximation of x on to w and what is the error? The error is the remaining part of x which is x w perp. This is the error vector and therefore, the length is the square. is the square So, the error is given by the orthogonal length of the orthogonal complement and the approximant is the original vector.

(Refer Slide Time: 23:22)



Let us, look at one simple example; let us take R 3. So, in R 3 consider the subspace which we are seen before the subspace consisting of all vectors a a a that means these are vectors all of those components are equal. First component is a, second component is a, and third component is a. we have this subspace already seen that this is a subspace. Now suppose, we take let us x be any vector in R 3. So, we want to find best approximation of x from W. What we are learnt so far says that in order to find the best approximation, you have to look at the orthogonal projection. So, we know this is given by the orthogonal projection of x onto W. we have seen in the earlier lectures that the orthogonal complement in this situation is exactly given by we have seen these early lectures. previous lecture that the earlier lectures

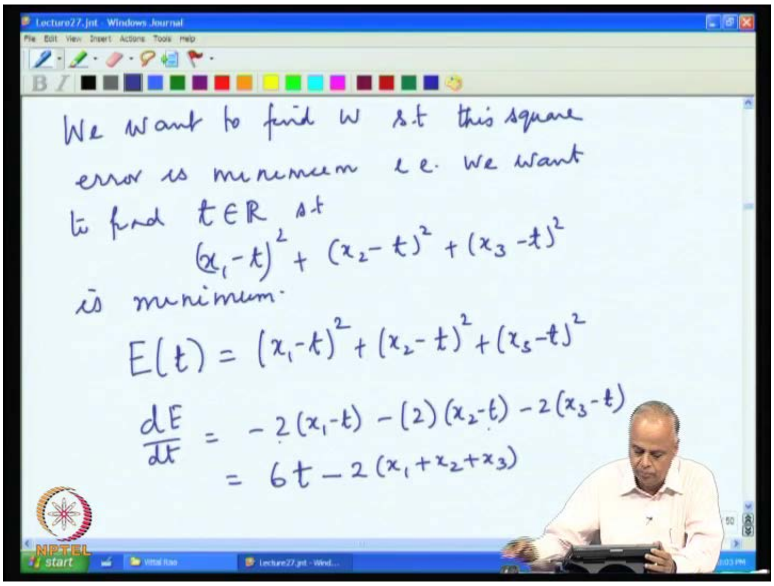
(Refer Slide Time: 25:24)



We have see that the orthogonal projection x w is precisely x 1 plus x 2 plus x 3 by 3 x 1 plus x 2 plus x 3 by 3 x 1 plus x 2 plus x 3 by 3. So, the vector which approximates x best from w is given by taking the average of all the components. So, we have the vector x is equal to x 1 x 2 x 3 w wants all components to be equal. So, we want to construct a vector from x where all components are equal with least error.

The best thing you should take all the components should be the average of the components o f x when we have x 1 plus x 2 plus x 3 by 3 x 1 plus x 2 plus x 3 by 3 and so on. Now, let us verify this directly let us verify this directly what do we mean by this? Take any w in W, now the any vector in w has all components equal. So, let w equal to t t t where t is a real number. Now what is the square error? The square error is the difference times the length now what are the the length of the vector, what are the difference vector x is the vector x 1 x 2 x 3, why the w is the vector t t t. So, it is going to be equal to x 1 minus t x 2 minus t x 3 minus t. This vector square, but the length of the vector square is the sum of the components squared. So, x 1 minus t squared plus x 2 minus t squared plus x 3 minus t squared. Now, we want to find that vector which minimizes the error, that means we want to find the value of t for which this value, this is minimal.

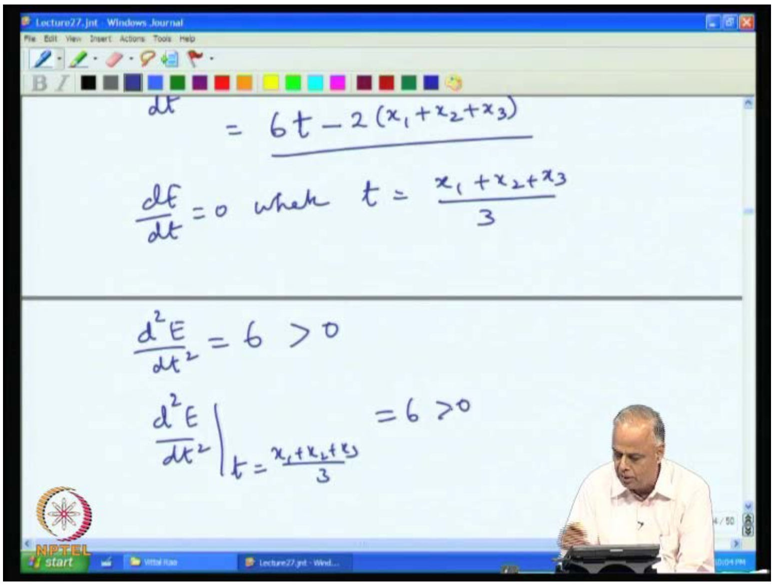
(Refer Slide Time: 27:42)



So, we want to find w such that this square error is minimum. That is we want to find t in R such that the error square was x 1 minus t squared x 2 minus t whole squared x 3 minus t whole squared is minimum. We are given the vector x and therefore, x 1 x 2 x 3 are fixed. So, we add to click those a function of t we have a function of t and we have trying to find the minimum in calculus, we are learned that to find the minimum, first find the places we have the derivatives vanish, that is call the critical points and this side it at a critical point the second derivative is negative, it is maximum and it is second derivative is positive, it is a point of minimum.

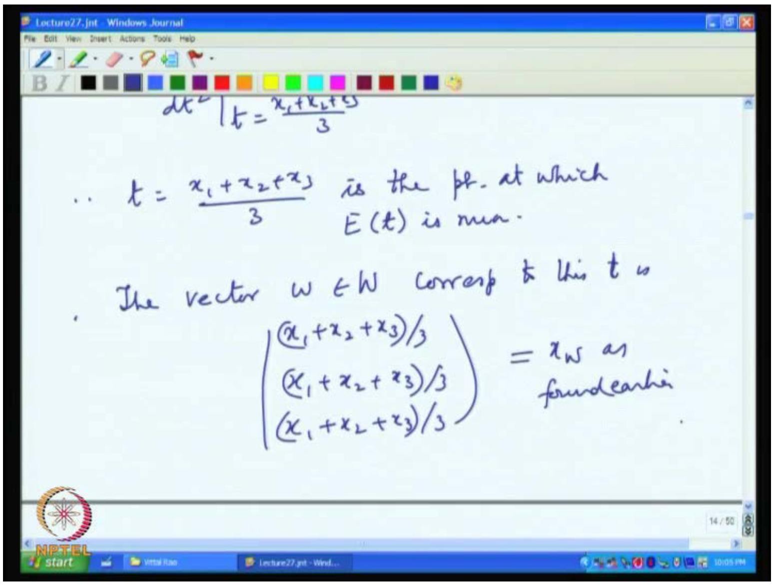
So, let us called this as the E t, the error function depending on t what vector you choose, so this is e t and we have trying to minimize this function .Therefore, we differentiate this function with respect to t and we get minus 2 x 1 minus t minus 2 x 2 minus t minus 2 x 3 minus t, which is equal to there is a 2 t, there is a 2 t, there is a 2 t so, there is a 6 t minus twice x 1 plus x 2 plus x 3.

(Refer Slide Time: 29:56)



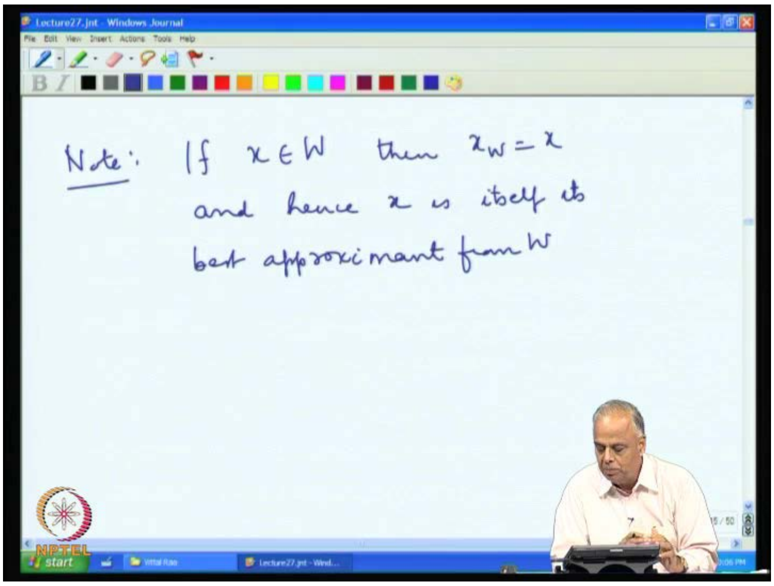
And therefore, the derivative is 0, when t is equal to x 1 plus x 2 plus x 3 by 3. This equal to 0 look at t is equal to x 1 plus x 2 plus x 3. So, there is only 1 critical point which is t equal to x 1 plus x 2 plus x 3 and if you see that d squared e by d t squared is just 6. It is always positive and therefore, if you have a critical point and secondary positive and therefore, this is always six and therefore, d squared e by d t squared, when t is equal to x 1 plus x 2 plus x 3 by 3 is also 6 greater than 0.

(Refer Slide Time: 30:45)



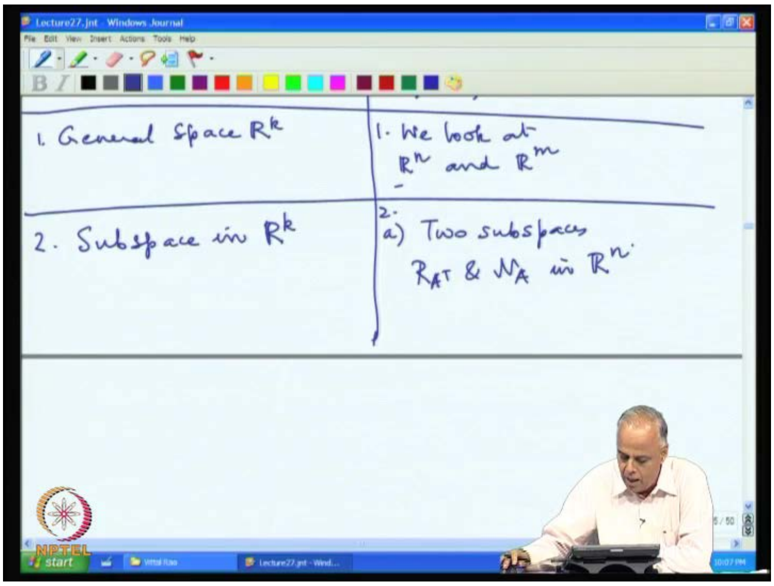
And therefore, t is equal to x 1 plus x 2 plus x 3 by 3 is the point at which e t is minimum. And the vector corresponding to the therefore, the vector w in w corresponding to this t is x 1 plus x 2 plus x 3 by 3, x 1 plus x 2 plus x 3 by 3 because t t t all components are t, now a form t value as. This was equal to x w as forward so thus we find that the best approximation is indeed if indeed the projection of x onto w.

(Refer Slide Time: 32:00)



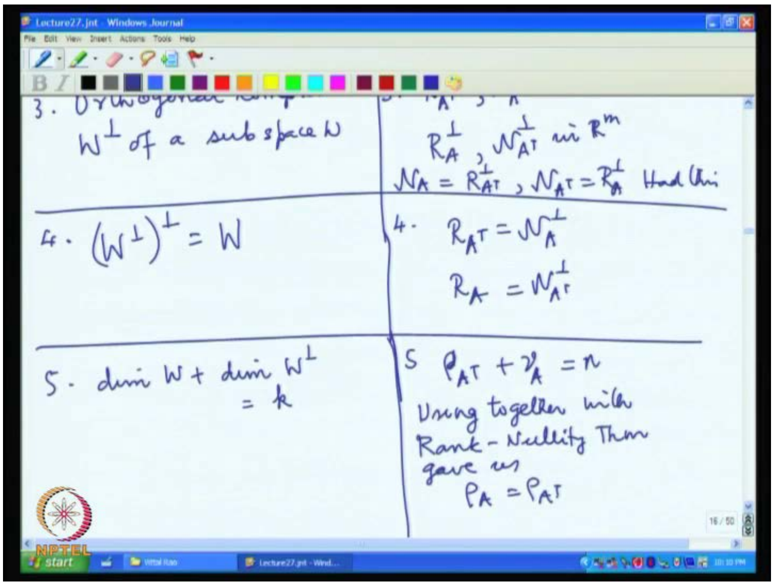
So, thus now in know that we know how to approximate a vector from a subspace, note if x is in w then he is himself its best approximation if project x on to w, you get x then x w equal to x and hence, x is itself its best approximate on w naturally because already in w. Now, your learn the lot of things about a subspace is orthogonal complement best approximation etcetera. Let us now consolidate all this and see how we are interpreted them in the context of matrices.

(Refer Slide Time: 32:50)



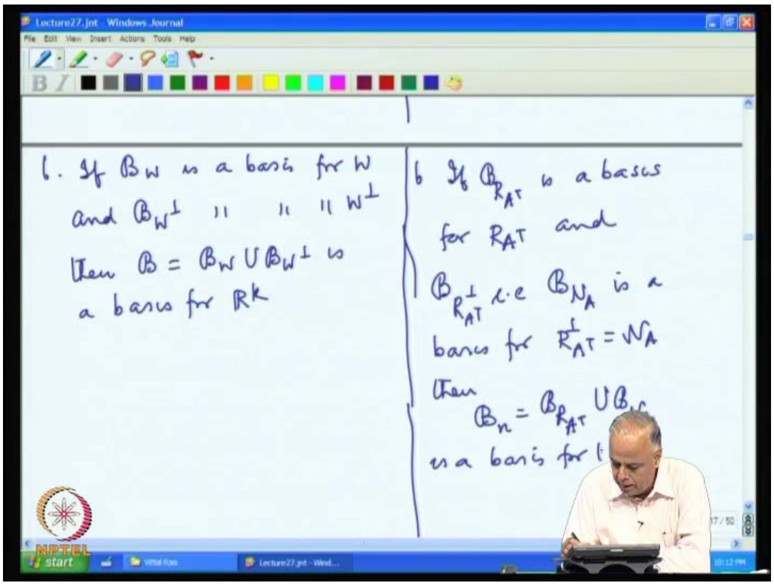
First we will divide in 2 parts, 1 in general vector spaces R k and then we look at the matrix context. How we use these in the matrix context. So, we take a matrix which is in m by n matrix. We have we look at the general space R k in the matrix context we are look at 2 subspaces R n and R m, because the matrix a converts R n vectors into R m vectors. So, we look at R n and R m. The 2 spaces like the R k, which are of the interests on the matrix situation, then we introduce the notion of a subspace in R k now in the matrix context we have 2 situation again R n and R m we had 2 subspaces in R n 2 subspaces in R m, we have 2 subspaces the range of a transpose and null space of a in R n and b the 2 subspaces the range of a and the null space of a transpose in R m.

(Refer Slide Time: 34:40)

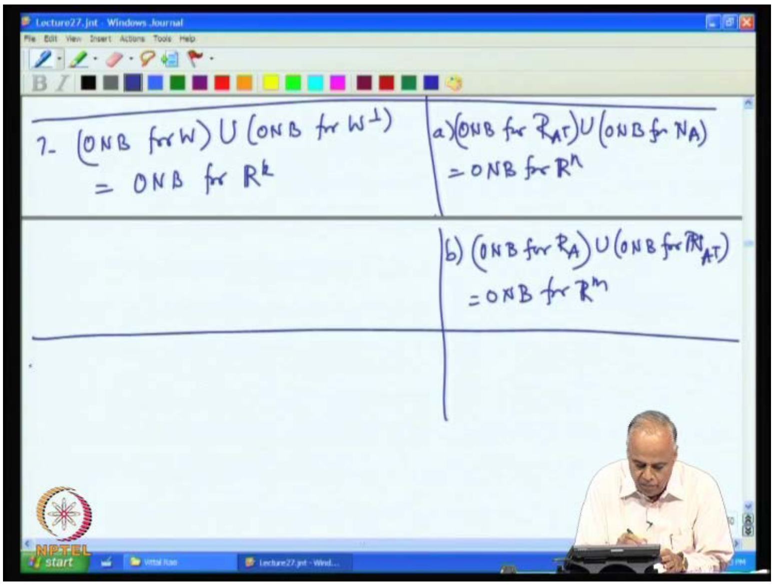


That is the second thing that we observe the notion of subspaces how will use in the matrix context then we had the notion of orthogonal complement W perp of a subspace W. We use this on the matrix context for these four subspaces that we have so we have R A transpose perp and N A perp in R N and then we have R A perp and n a transpose perp in R M. When we did that we had a property of the that we observe n a was R N transpose perpendicular and N A transpose was R A perp we had this, then we that the perp of is the original the orthogonal complement of the orthogonal complement is the original subspace w, when use that in the matrix situation, we got that the R A transpose is N A perp and R A is N A transpose perp. Consequently using this along then we had dimension of w plus dimension of w perp was equal to k when we use that here the rank of a transpose plus the null space nullity of a is equal to n but this plus use in together with using together with rank nullity theorem gave as rho a equal to rho a transpose that is the dimension the rank of any matrix is equal to the rank is transpose.

(Refer Slide Time: 37:06)

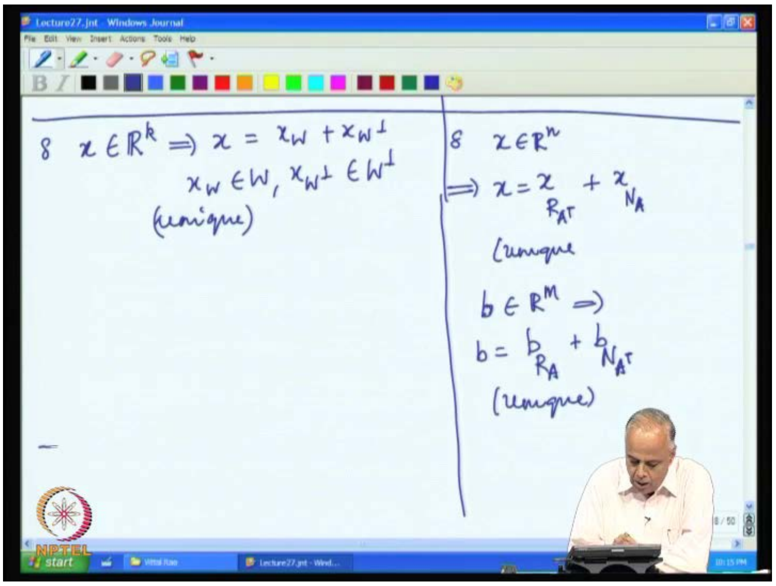


We further have that if B is a basis for w and B w perp is a basis for w perp then B equal to b w perp union b w perp is a basis for R k. If you use this notion in the 2 r n and r m spaces for the matrix we get if B R A transpose is a basis for range of a transpose and b r a transpose perp, but what are the R A transpose perp which was same as B N A, so we will put it the first. That is b N A is a basis for r a transpose r a transpose perp which is equal to N A then b n equal to B R A transpose union B N A is a basis for r n. Similarly and similarly, the basis for r a union a basis for N A transpose equal to b a basis for r n, let we put it as b m so we have got this basis concept and correspondingly orthonormal basis. (Refer Slide Time: 39:10)



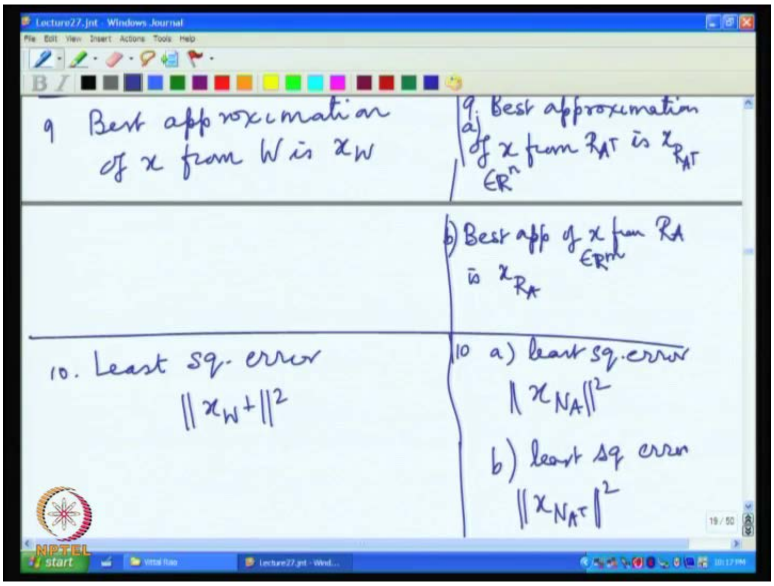
So, we have orthonormal basis for w union orthonormal basis for w perp is equal to orthonormal basis for R k. If you now apply these to matrix situations orthonormal basis for range of a transpose union orthonormal basis for range of a transpose per which is N A is equal to orthonormal basis for R N and further we have orthonormal basis for range of a union orthonormal basis for range of null space of a transpose is equal to orthonormal basis for r m. So, these notions of basis orthonormal basis decomposition orthogonal, decomposition etcetera.

(Refer Slide Time: 40:19)



We have again found that x belongs to R k implies x is equal to x w plus x w perp x w belongs to w x w belong perp belongs to w perp and this decomposition is unique, any vector in R k can be decomposed as the sum of a vector in w and a vector in w perp. Now, we consolidate this here on the r n side we get x belongs to r n x is equal to we take w to be x in R A transpose plus x in R A transpose perp r a transpose perp is N A and these decomposition is unique and on the R m side b belongs to R m that implies b can be transpose as decomposed as a part which is in r a and take in w as R A, now plus a part which is in R A perp but R A perp N A transpose. Any vector in the n dimensional end dimensional side can be decomposed as a vector in R A transpose and vector in null space of a and any vector in the m dimensional space can be decomposed as a vector in R A and a vector in N A transpose. These decomposition are unique and then finally today we learn that the best approximation,

(Refer Slide Time: 42:08)

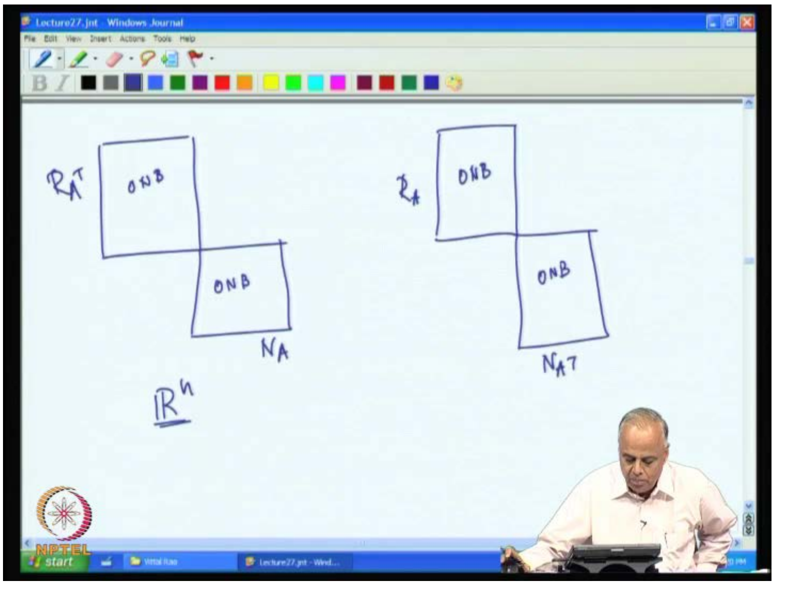


So the best approximation of x from w is x w, so if you now use this idea in the matrix situation subspaces implies of w will is r a transpose once and r a on the other side once. The best approximation of x from R A transpose is the projection of x of R A transpose best approximation of x from R A is the projection of x R A, and this x receive the R m side and this is the R N side. So, any vector in r n can be approximated with least square error from r a transpose and any vector in R m can be approximated with least square error in R N. Then, what is the least error is the norm of x w perp square.

So, in this in the a case that is approximating vectors in R N by vectors in r a transpose, the least square error is the projection of x into N A and that because R A perp R A transpose perp is N A and similarly, in the R m side the least square error is norm of x N A transpose square. So, all these notions we are studied so far have the role to play in the matrix situation with respect to the four subspaces that we are since we repeatedly asset the that these 4 subspaces are going to play a very major part in our analysis.

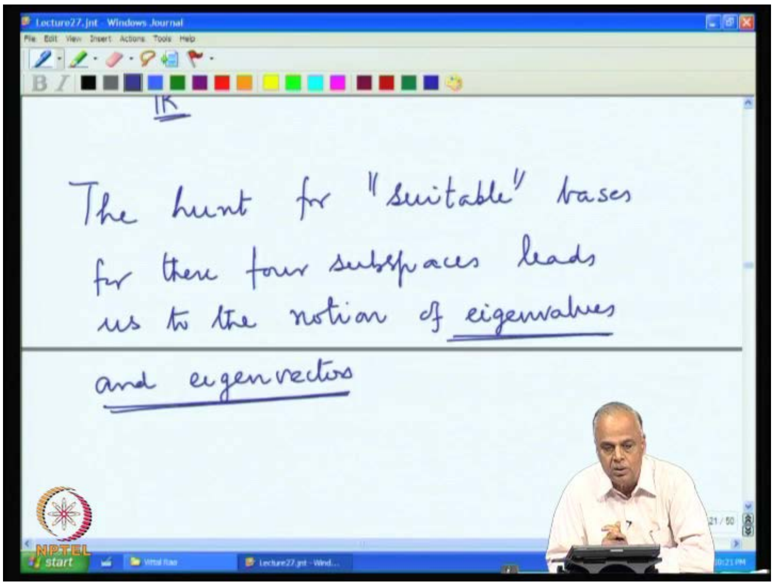
We are consolidated a lot of information about these 4 subspaces out of deal with them we are realize the district r n for example, is split into 2 parts but by R A transpose and N A and R m is split into 2 parts by R A and N A transpose and we are seen that the vector can displayed they can be approximated with least error from both these spaces and all these fact will be use the basis can be put together to basis orthonormal common basis can be put together to common orthonormal basis and so on. So, these are the facts that will be repeatedly using now, the crucial part in all these is this the orthonormal basis we seen that the orthonormal basis for w and the orthonormal basis for w perp together gives orthonormal basis for R k similarly, therefore if you now, take the main part again a repeat the main part

(Refer Slide Time: 45:46)



Therefore, is if you now choose R n and we look at the matrix which gives the decomposition, we have this decomposition the R n is split in to r a transpose and N A and the r m is split in to the range of a and N A transpose. This splitting is an orthogonal splitting and therefore, any vector can be decomposing and these we are guessing before. So, the what we are looking for in the fact if you now, choose an orthonormal basis for this and an orthonormal basis for this the 2 together their union will give me a orthonormal basis for R m then an orthonormal basis for this and an orthonormal basis for this, the 2 together will give me a basis for r m. So, our main aim is to find this four orthonormal basis the orthonormal basis can be found but we want to find them in such a way we want to link them in such a way that we using this linked basis we can asset the answers for all these questions about the matrices that we reset, so the real strategy lies in the strategy for choosing this basis.

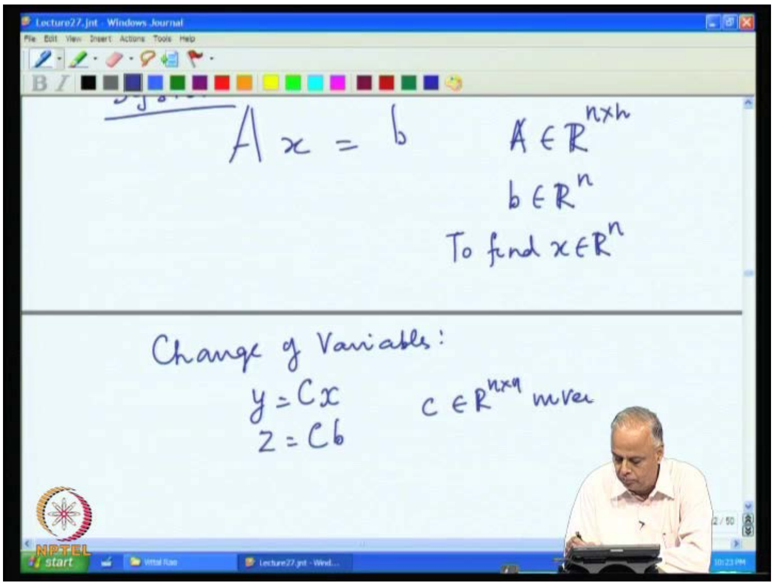
(Refer Slide Time: 47:22)



We shall look at this strategy as, we go along the main this which the hunt for suitable basis, we will explain as go along what is suitable for as we look at the questions we ask and we will see, which is the basis that makes the burking easy for us gives the answer in easy language and basis that are suitable for us. So, the un for suitable bases for these four subspaces leads us to the notion of Eigen values and Eigen vector and therefore, this notion of Eigen values and Eigen vectors is the very important notion in matrix stage and this as these role to play in the analysis of these structure of the matrix and the analysis of finding the various questions of the matrices. We are we will soon see, how they play this role but before we do that we shall look at another problem in which these matrices these notions of Eigen values and Eigen vectors play.

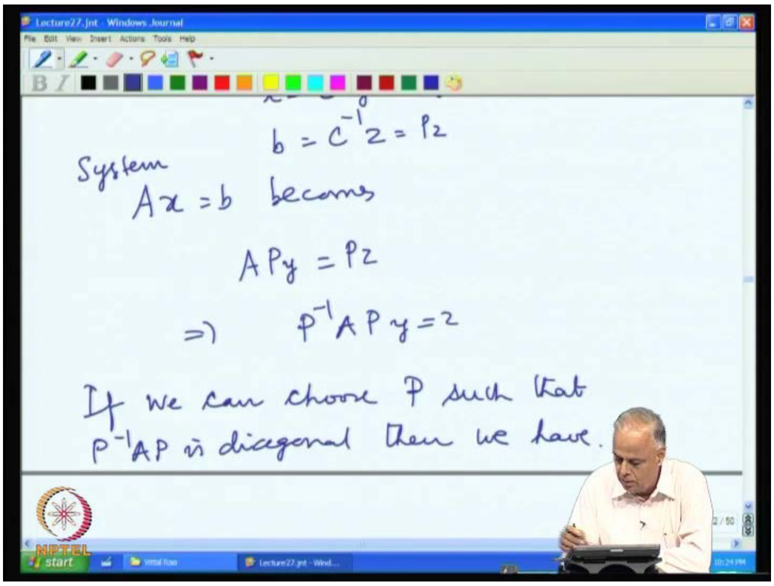
So, another problem where Eigen values and Eigen vectors notions these the main problems in which did appear from which we move onto the other question raised above notion appear. What is these problem now, let us again and again we should remember that the origin of linear algebra was in the analysis of system of equation linear systems of equation and in the beginning we said we shall first studies simple system among the simple system systems n equation with n unknowns where the simpler once.

(Refer Slide Time: 49:51)



So, what we did was we looked at a system a x equal to b system a was an n by n matrix and b was R n vector, we want to find x in on what we said was we want change of variables, because we said that diagonal systems are easy, we want to change variables to bring it a diagonal system so, we said y equal to some changed x; z equal to some changed b, where c is an n by n matrix and m vertible, when we change this variable.

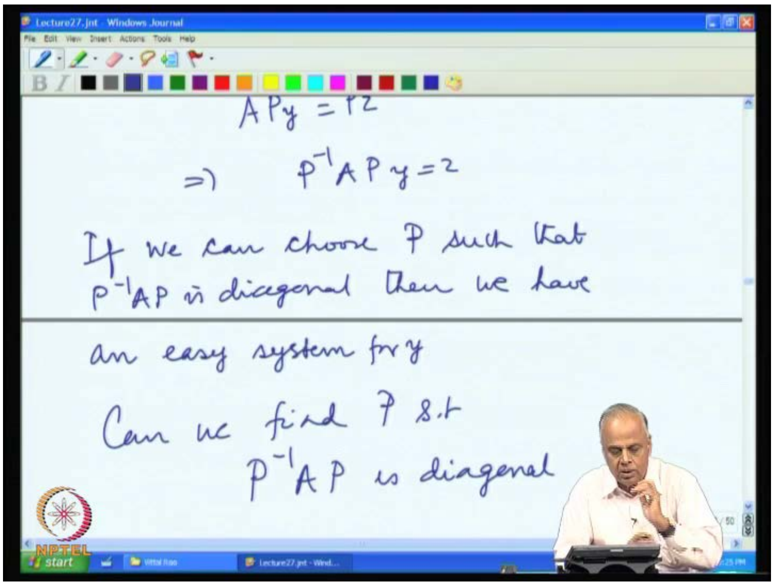
(Refer Slide Time: 50:40)



We can write x as c inverse y if which we call as p y, b as c inverse z which we call as p z, then if you look at the system A x equal to b. Now, with the change variables becomes

A x equal to R, p inverse A p y equal to z. If we can choose P such that p inverse A p is diagonal then, we have an easy system what diagonal system for y and therefore, we can analyse y and once we analyse why we know the information about x.

(Refer Slide Time: 51:51)



So, therefore the main question is can we find a P such so, can we find p such that P inverse a p is diagonal. This is the main problem which leads as to the notion of Eigen values and Eigen vectors. We shall see in the next lecture how such a question leads as to the notion of Eigen values and Eigen vectors and how this notion of Eigen values and Eigen vectors play crucial role in the answer to these question and how these Eigen values and Eigen vectors appears in the formation of these hunting for these 4 suitable basis.