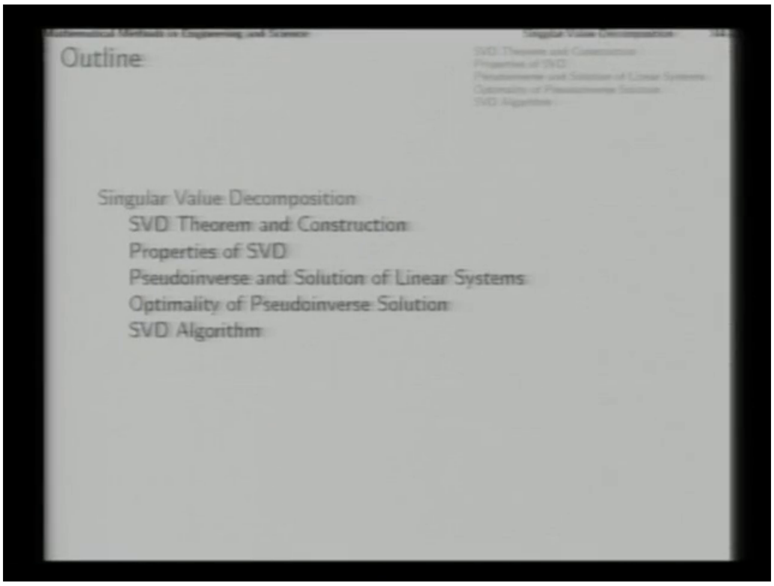
Mathematical Methods in Engineering and Science
Prof. Bhaskar Dasgupta
Department of Mechanical Engineering
Indian Institute of Technology, Kanpur

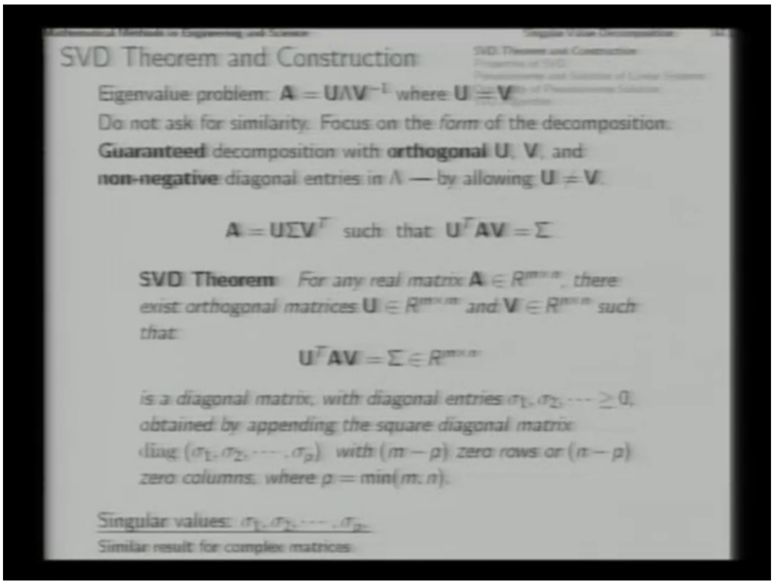
Module – III
Selected Topics in Linear Algebra and calculus
Lecture – 01
Singular Value Decomposition

Good morning, in this lecture, we will be studying singular value decomposition this topic embodies a very deep connection between quite a few different topics in the area of linear algebra.

(Refer Slide Time: 00:24)

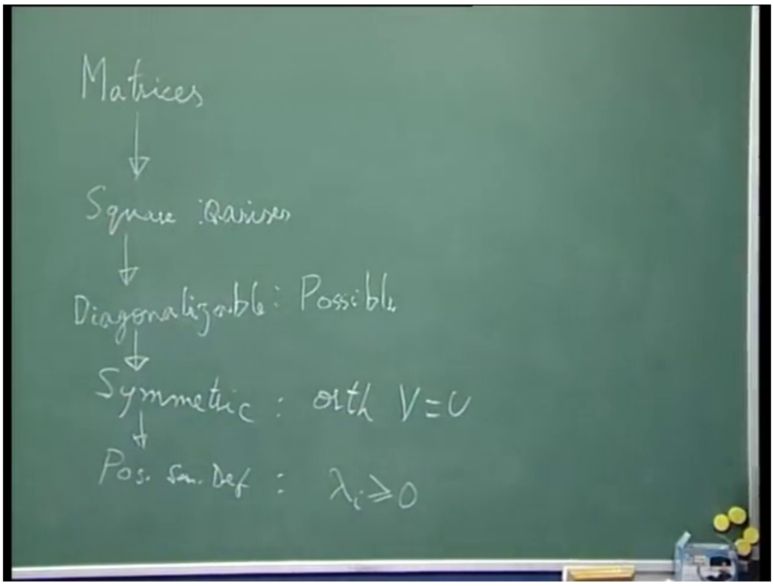


(Refer Slide Time: 00:37)



Consider this situation we have already studied Eigenvalue problem in which we wanted to decompose a matrix A in this form with U and V equal, we have already studied Eigenvalue problem and all the time of our study in Eigenvalue problem we have faced this question whether the decomposition of this sort will exist or not if it exists, then how to handle it and so on. So, it would be nice always in the Eigenvalue problem; if we could make this lambda diagonal with U and V orthogonal and also such things and at every step our work was made with difficulties of several sorts first among all matrices.

(Refer Slide Time: 01:36)



We could ask this question only for those matrices which are square that is a sub set a sub set of matrices the sub set of square matrices constitute the only matrices for which this question arises.

So, for square matrices this question arises to begin with even in the square matrices not all matrices can be diagonalized not all square matrices can be diagonalized. So, among square matrices we had a subset which is the set of diagonalizable matrices for which this is this kind of a decomposition is possible, right.

Even among diagonalizable matrices we had another sub set of matrices which are symmetric for which this decomposition would be affected with orthogonal V right which is same as U for that matter with that condition full filled even among the symmetric matrices for which we had this valuable theorem that you can work out an orthogonal digonalization even there the diagonal elements of lambda could be negative now even among symmetric matrices we had a sub case sub set which is positive semi definite in which case the lambda i turns out to be non negative. Now this is the best possible situation which we could sink of and that is a sub case of the sub case of the sub case of the generalized general form of the matrices.

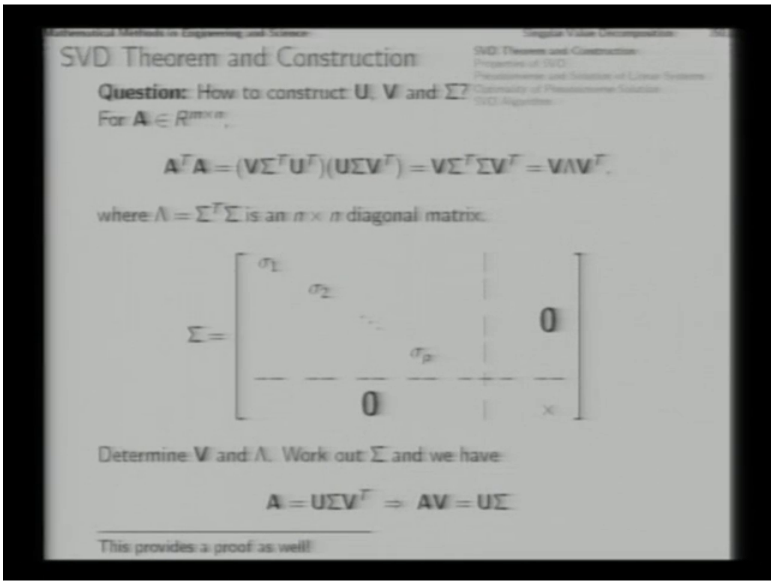
Now, we can ask this question that we do not ask for a similarity transformation and we focus on this form of the decomposition when you say we do not ask for similarity we basically want to allow U and V to be different. So, in that case, we ask this question that if we do not ask for U and V to be equal then what are these we can ask for and get results and with just this one relaxation of allowing U and V to be different; different in content as well as in size if we allow that then we can get a decomposition of this sort which is guaranteed for all matrices irrespective of size and shape; that means, even a rectangular matrices with orthogonal U and V matrices and with non negative diagonal entries in delta in this matrix lambda diagonal matrix lambda that is in that case we do not refer to it as lambda because lambda has been already used for the matrix of Eigenvalues.

So, we show that as sigma; that means, that just by allowing this U and V to be different we can effect a decomposition of this sort with all the other phase sets that is the decomposition will be possible for all matrices and it will be always possible the question will arise for all matrices including rectangular that diagonalized that decomposition we cannot call it diagonalization that decomposition will be always possible with orthogonal U and V not same anymore and the diagonal entries of this matrix sigma will be all non negative such a decomposition is the singular value decomposition and those diagonal entries are called singular values of the matrix A underline is this very important theorem called the SVD theorem or singular value decomposition theorem.

The theorem says for any real matrix A of size m by n there exists orthogonal matrices U which is m by m and V which is n by n both orthogonal such that U transpose AV is diagonal matrix of size m by n. Now, what is this idea of a diagonal matrix of a rectangular size? So, its diagonal entries are sigma 1, sigma 2, sigma 3, etcetera, all non negative which you obtain by getting f square matrix first of size p by p in which p is lesser of the 2 dimensions m and n.

Now, if you want this diagonal matrix to be m by n size, then whichever is larger m or n that many extra rows below rows or that many extra 0 columns you append and these diagonal entries sigma 1 to sigma p are called the singular values of this matrix A. Similar result is there for complex matrices then for that the as many theorem will read for any complex matrix A belonging to c m by n they are exists unitary matrices U and V such that U star AV where star is a conjugate transpose is real sigma this is always real and so on. So, now, this theorem gives the basis for the decomposition in this manner for a matrix A.

(Refer Slide Time: 08:05)



Now, the question arises how to construct U V and sigma the 3 components the 3 factors the way we work out their construction at the same time provides a proof also of the SVD theorem that such factors U V sigma will always exists. So, let us quickly look at the construction. So, construct the singular value decomposition the factors U sigma and V you first say that if we would decompose a in this manner A as U sigma V transpose then it transpose a transpose will be this V sigma transpose U transpose and then we can just multiply it as we multiply it; U being orthogonal U transpose U will be identity and we have got V sigma transpose sigma V transpose; now sigma transpose sigma.

We have already discussed that sigma is a matrix of this shape in which if m is less then it will have only m columns which will mean that the matrix will be only this much these rows will not be there since m is less, then it will have this shape if n is less, then it will have this shape. So, extra 0 rows or extra 0 columns there will be no question of anything here because one of these 0 blocks will be here not both.

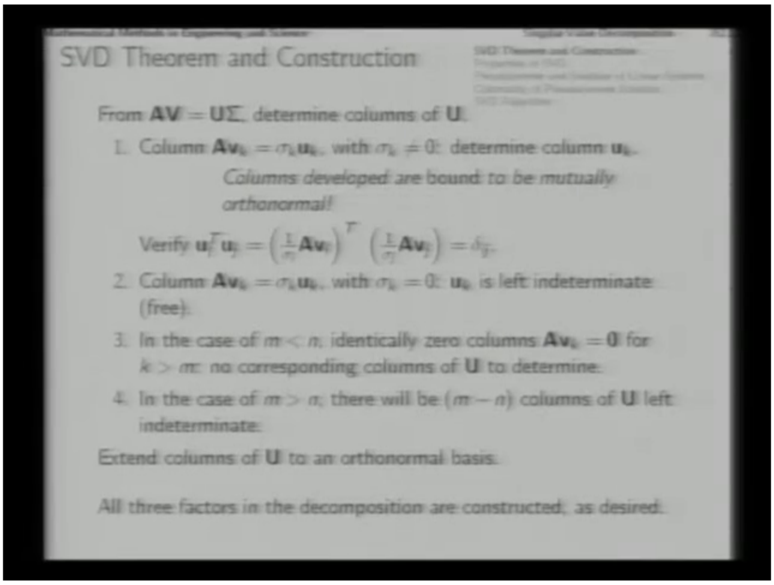
So, if sigma is of this shape then sigma transpose sigma will be a square matrix in which the diagonal entries will be sigma 1 square sigma 2 square up to sigma p square and then since this matrix is n; n size. So, if n is larger than there will additional 0 entries in the rest of the diagonal position and all the of diagonal entries will be 0. So, that is the description of this sigma transpose sigma now here this sigma transpose sigma this matrix is being called lambda which has a reason you see a transpose a is certainly symmetric not only symmetric its positive a semi definite also you cannot say a priori whether it is positive definite or not, but positive semi definite it is certainly be.

Now, if this is a symmetric matrix then this certainly has a diagonalization and orthogonal diagonalization for that matter and this V lambda V transpose is actually the decomposition that you do when you solve the diagonalization problem of a symmetric matrix so; that means, this V which you want in singular value decomposition is in fact, the matrix I can vectors of a transpose a and this lambda then is the diagonal matrix of Eigenvalues of a transpose A if so, then we already know how to determine V n lambda because we have studied the Eigenvalue problem of a symmetric matrix in good detail, we can effect this diagonalization so; that means, by effecting the diagonalization of a symmetric matrix we determine V and lambda the moment V and lambda determined we can work out sigma because each diagonal entry of lambda the first png s are nothing, but sigma 1 square, sigma 2 square, sigma 3 squared and up to sigma p squared, right. So, from the first p lambdas from here which are all non negatives we can take the square root. So, when you take the square root there are 2 square roots.

For a positive number 1 positive 1 negative. So, you collect only the positive ones which you put as sigma 1, sigma 2, sigma 3, etcetera, up to sigma p. So, all the non trivial entries of this matrix sigma as sigma 1, sigma 2, sigma 3, etcetera, up to sigma p. So, all the non trivial entries of this matrix sigma is now our hand then and then we append that with appropriate number of 0 rows or 0 columns depending upon what is the size of a which is the same as the size of sigma; that means, V and sigma are now in our hand.

Now, remember a is U sigma V transpose and V is orthogonal so; that means, we can post multiply that original definition of the singular value decomposition with V and then we de transpose V will identity from here you will get only U sigma and on this side you will get a b in which in this entire equation a was originally given V and sigma we have determined and we are left with the problem of determining this matrix U the columns of the matrix U.

(Refer Slide Time: 13:14)



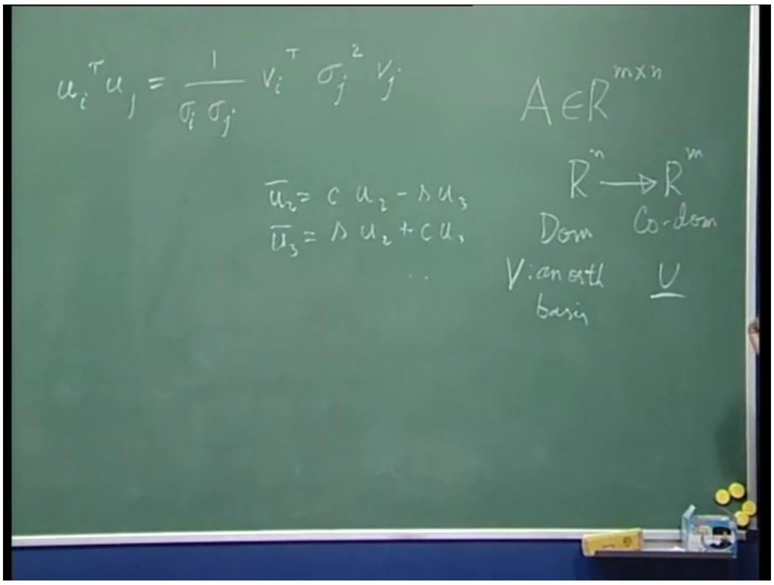
So, 4 situations will arise when we go to determine the columns of matrix U. In fact, 4 situations may arise in any particular case only 3 of them will arise the there either the third will arise or the fourth will arise depending upon whether the matrix A has more rows or more columns. So, first situation is actually the one in which you will have some information to determine if you equate the 2 sides column by column, then you will find that the left side will give you columns which is Av 1, Av 2, Av 3 where v 1, v 2, v 3 are columns of matrix V and from the right side you will get you will get columns the corresponding columns as U 1 into sigma 1 plus all 0s, then U 2 into sigma 2, then plus all 0s and so on; that means, you will get this kind of column equations when you break this column by column that will be the first R columns if R is the rectangle; that means, for the Nonzero singular values.

So, out of these p singular values some of them may be 0, right. So, for Nonzero singular values corresponding column equations will give you this kind of equations and if sigma k is Nonzero the determining the corresponding columns of U is easy you just divided AV k by sigma k and you get the columns of U. So, these columns developed from here are bound to be mutually orthogonal.

You can verify that suppose 2 columns U y and U j have been developed like this and you want to find out U y transpose U j. So, they are not only orthogonal they are orthonormal that is each of them is a unit vector also. So, being orthonormal this has to be 1 if i and j is same and 0 if i and j is i and j are different. So, you can see this that when you consider U y transpose U j from these expressions from here you have determined U y U j, then here you will find that you will get V i transpose a transpose AV j. Now a transpose A is the matrix for which we actually solve the Eigenvalue problem, right. So, V j is its Eigen vector corresponding to Eigen value lambda j that is sigma j square.

So, when you write this here 1 by sigma i is here 1 by sigma j is here, right.

(Refer Slide Time: 16:11)



So, we collect the scales together and then we are left with vi transpose vi transpose a transpose avj. So, write vi transpose and a transpose a vj is lambda j vj this is lambda j this is vj. So, lambda j that is sigma j square is scalar which we can bring here and we are then left with v i transpose v j here from there you find that if i and j are different then vi transpose vj is 0 because v and lambda together give the orthogonal diagonalization of a that means.

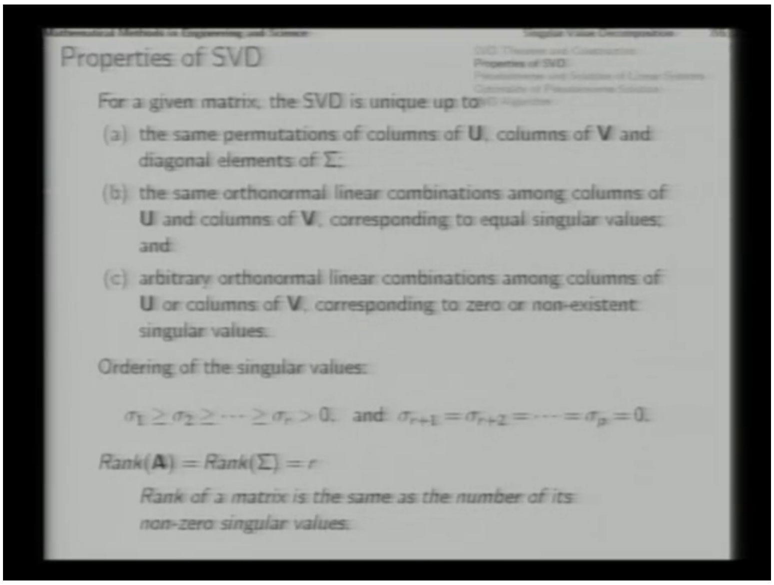
Columns of v are mutually orthogonal right. So, if i and j are different then v i transpose v j is 0 and you have got the orthogonally of u i transpose u j right here on the other hand if i and j are same then vi transpose vj transpose vj you will get which is 1 because v is orthogonal. So, each column v j in particular is of size 1. So, in that case v j transpose v j will be 1 and this sigma j transpose square cancels with this square i is equal to j in this case. So, you will get 1 here; that means, u j transpose u j will be 1. So, that shows the orthonormality of all the columns that we have determined from this; this much for those singular values which are not 0 right for Nonzero singular value.

For the singular values which are 0 we have got this AV k equal to sigma k u k and sigma k is 0; that means, you are talking about AV k equal to 0 right. So, the corresponding u k is left in determinant so; that means, that you cannot determine uk from this relationship because the coefficient is 0, but it is left in determinate; that means, you are free to chose a suitable uk what is a suitable uk a unit vector that is orthogonal to all the other columns that we have already determined right now in a case where m is less than n; that means, u has less number of columns and v has more number of columns right; that means, in that case you will get further equations AV k for k greater than m for which on this side you will get 0s right and from that there is no corresponding column of u to determine right.

So, this is gone the fourth case is where m is greater that is the matrix A has more rows than columns in that case after all this calculations there will be further row columns of u which are left indeterminate. So, just like the case 2 in this case also there are additional 2 columns additional columns of u which are left indeterminate. So, just this case in this case also the additional u vectors are determined to make the entire u matrix orthogonal; that means, additional columns of this case with 0 singular values and additional u columns corresponding to this case with additional singular values additional columns which have no matching.

Singular values. So, these 2 cases are determined based on the orthogonality requirement of u so; that means, in one line you can say we extend the columns of u determined from here to an orthonormal basis and that full set of m vectors will give you the square matrix u. So, this way after the 3 factors of the singular value decomposition have been constructed you have a equal to u sigma b transpose each other c you have in hand after constructing the singular value decomposition like this you would like to see what are the properties of such a decomposition. So, first question after verifying existence is uniqueness is it unique.

(Refer Slide Time: 20:53)



The actual answer is that it is actually not unique for example, you can apply several changes in it and still the changed U sigma V will constitute another singular value decomposition of the same matrix so; that means, that you can do several changes.

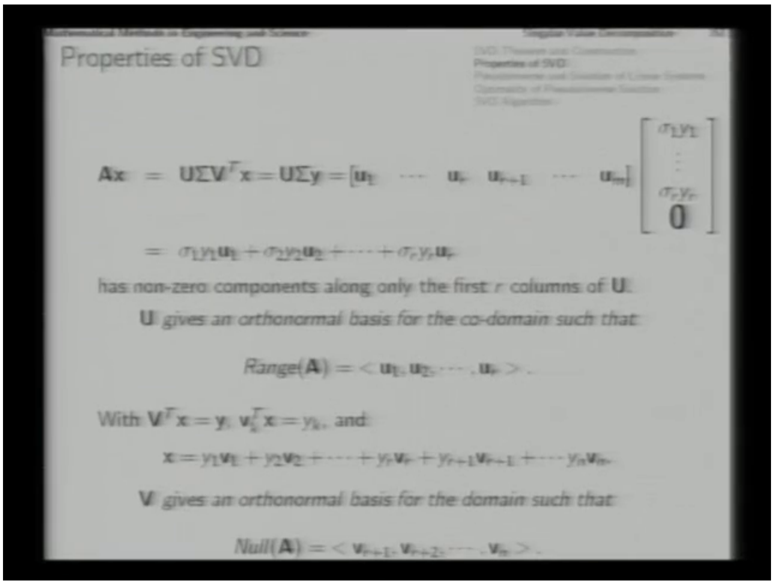
So, those changes are here and then you can say for a given matrix the SVD is unique up to these changes; that means, it is actually not unique it is determinate, but such changes will not disturb the requirements such changes will not disturb the fact that the matrices the decomposition is still and SVD of the given matrix. So, what are these changes which are possible the same permutation of columns of U columns of V and diagonal elements of sigma; that means, if you interchange sigma 2 and sigma 5 and at the same time interchange columns U 2 and U 5 and interchange v 2 and V 5 then the resulting U sigma and V will still give as SVD and so on.

Now corresponding to equal singular values you have got columns of U and v. So, among them if you work out an orthogonal re organization that is suppose sigma 2 and sigma 3 are same then you say that I will work out this now this will be my new U 2 and this will be my new U 3 and corresponding for V also between v 2 and v 3 also you will make the same transformation this will be still the resulting you and V matrices with the same sigma will still give you a singular value decomposition which is valid the particular case particular transformation that we worked out here is cos theta sin theta sin theta cos theta here that is cos theta minus sin theta note that this is minus. So, that matrix is an orthogonal matrix.

So, such orthogonal linear combinations for columns of U and corresponding columns of V is fine that will not disturb the singular value composition for 0 or nonexistent singular values you can do any linear combination any arbitrary orthonormal linear combinations among the columns of U or columns of V. So, that will still be alright. So, these reorganizations in an already existing SVD can be done and the result will be still an SVD.

Now, if this can be done, then we can do something better than what we have done till now that is we have determined sigma 1 sigma 2 sigma 3. Now if the permutations can be appropriate in that then we can order them that is we can organize columns of U and V in such a manner that the sigma the singular value comes first is the largest magnitude and. So, on this we can do. So, this is typically done when we work with singular value decomposition so; that means, the Nonzero singular values come at the top with this order and after that the 0 singular values come and after that of course, additional rows or columns may come depending up on the rectangular size and shape of the given matrix right now here what is R? R is the rank and this is a very simple result which you can immediately establish that is rank of the given matrix is a same as rank of sigma which is R here other properties.

(Refer Slide Time: 24:58)



You would have already noticed that matrix of matrix A is of size m by n; that means, it maps vectors from R n to R m right in which this is the domain and this is the co domain right now you can see that V being an n by n orthogonal matrix can give a basis which is orthonormal basis the columns of V are actually n dimensional vectors and they are all mutually orthonormal so; that means, that the columns of V give us an orthonormal basis for the domain similarly columns of U will give an orthonormal basis for the co domain and now here we see how these new basis V and U decompose the domain and codomain in to orthogonal sub phases.

So, you consider the application of a on a arbitrary vector x with a written as U sigma V transpose now if you represent the vectors in the domain the vector x in the domain in this new basis V, then the expression the coordinates of that those vectors in this new basis will be V transpose x actually V inverse x, but since V is orthogonal. So, it will be same as V transpose x right.

So, if we call that y then will have U sigma y U is written here and recognizing that sigma is a diagonal matrix which sigma 1, sigma 2, sigma t written on the diagonal entries among which the top R are Nonzero, you will have sigma y as sigma 1 y 1 sigma 2 y 2 etcetera up to sigma R y r below that everything else is 0 right and U has been broken and written in this fashion R columns here and then rest of them here now when you consider this product you will find the product is sigma 1 y 1 into U one plus sigma 2 y 2 into U 2 and so on up to this.

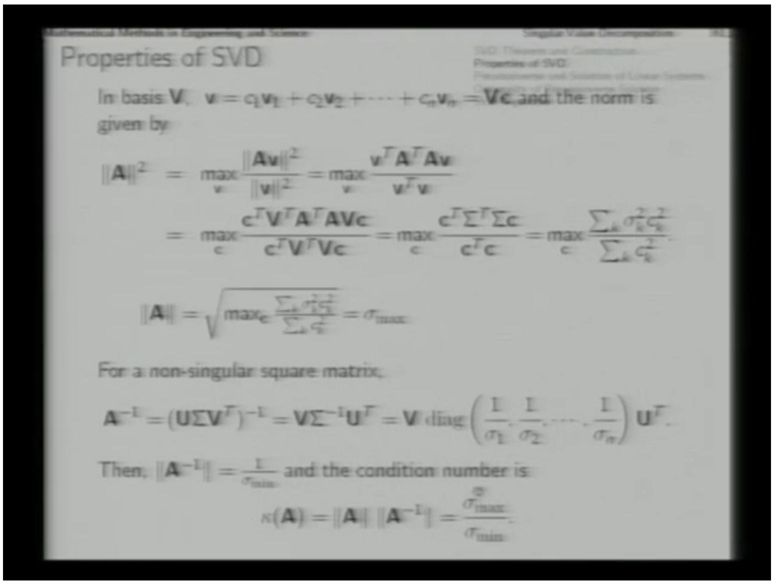
After that everything else is being here now see what is happening in this some you will notice that this has nonzero components along only the first R columns in this product the component along U R plus 1 U R plus 2 U R plus 3 etcetera are all 0; that means, that a x has nonzero components along only the first R columns of U right; that means, U has given as an orthonormal basis for the co-domain in which the range the vectors ax are contained only with the first R columns of u; that means, U gives an orthonormal basis for the co-domain such that the range is exactly described by the first R members of U and the rest of them described and orthogonal component of range orthogonal complement of range so; that means, the entire co-domain has been decomposed into 2 orthogonal subspaces the first one is the range which is x described with the first R columns of u.

Which are corresponding to the nonzero singular values and the rest of them are components in the orthogonal complement of range which are not in the range right similarly on the domain side if you see this V transpose x is y right. So, V transpose what are the rows of U transpose rows of V transpose are v 1 transpose v 2 transpose v 3 transpose and. So, on right and where v 1, v 2, v 3 are columns of v. So, the entries the coordinates in y; y 1, y 2, y 3 are actually v 1 transpose x, v 2 transpose x, etcetera. So, that is V k transpose x is y k that us the coordinate y k is found like this, right. So, that is component of x along the unit vector V k.

So, the full x is component of it along v 1 into the v 1 unit vector plus its component along v 2 into unit vector v 2 and so on like this now in this you will find that those vectors which are here only make a contribution in the ax mapping those here we will not make any such contributions because y r plus 1, yr plus 2 etcetera are 0s that we have already seen, right; they are made 0 by the in this product sigma y. So, whatever is y r plus 1 y r 1 plus 2, etcetera sigma y will kill their contributions; that means, whatever is y r plus 1 yr plus 2 etcetera their contribution in the product here will be 0 because sigma multiplied to them will kill their contributions so; that means, V here gives you an orthonormal basis for the domain.

Such that the components v r plus 1, v r plus V n they area actually constitute the null stage. So, you find that on the co-domain side range is constructed by the columns of U corresponding to nonzero singular values and on the domain side the null space is spend by the other columns other columns of V that is columns of V which are corresponding to the 0 singular values or non existence singular values and that is it. Now with this understanding in the background we proceed.

(Refer Slide Time: 31:40)

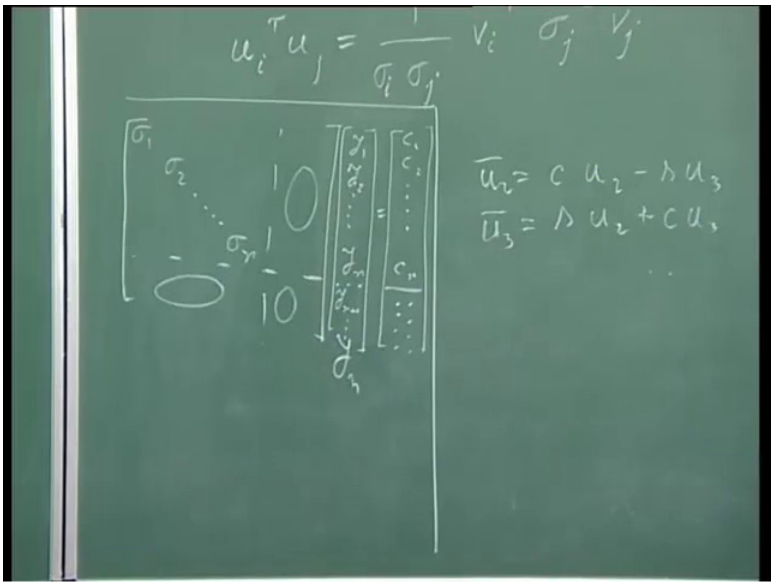


And find a few more interesting things in particular we work out the revised definitions of a normal matrix and the condition number of matrix in basis V if we write a vector in the domain in this manner then this can be written as V c right where V is the matrix with columns v 1, v 2, v 3, etcetera up to V n and c is the vector with this scalar components then from the definition of norm which we have seen earlier in the chapter 7 of the text book in an earlier lecture we discussed. So, from the definition of the normal matrix we say that norm square is maximum over V of norm Av square by norm V square now in this if we insert this description of the general vector V that is V c. So, then first of all from the norm definition we get this and there in place of small V we insert V c then we get this for V we have V c and for V transpose we have c transpose V transpose now here we have already seen that a transpose A.

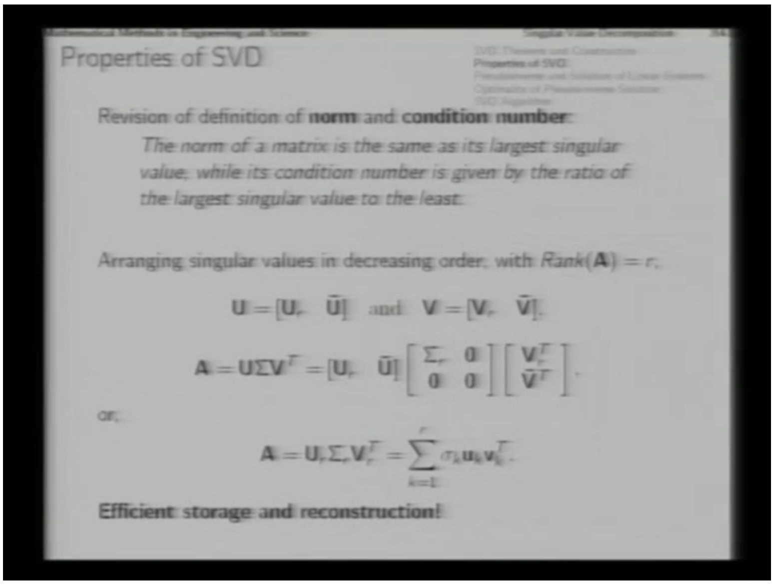
Diagonolazation was carried out with the basis matrix V and the corresponding diagonal matrix sigma transpose sigma right. So, in place of this whole thing we can write sigma transpose sigma right now here sigma transpose sigma diagonal matrix with entries sigma 1 square sigma 2 square up to sigma p square and then perhaps additional 0s right. So, this numerator based on to basically this right and now you say that we want the maximum of it when it will be maximum if sigma 1, sigma 2, sigma 3, sigma 4 are non all of the same magnitude then this is will be maximum when c k is a vector when c is a vector in which the only component is along the largest one which gets magnified by the largest amount then only you will get the maximum value of this and. So, you get the norm as norm square as the case where only that c k has a Nonzero value for which sigma k is maximum that is sigma max.

So, when you put sigma max there then you got this. So, norm is now found as the largest singular value of the matrix. So, this is the new revised definition of the norm of a matrix now for a non singular square matrix we worked out condition number right. So, here again we try to do that for a inverse we get this which is V sigma inverse U transpose which is this now you notice that by the same definition if we try to work out the norm of a inverse then it will be the largest singular value of a inverse and the smallest singular value from a will actually in its reciprocal will give the largest singular value for a inverse. So, you find that the norm of a inverse is 1 by sigma min of the original matrix a. So, the condition number is norm of a into norm of a inverse that is sigma max into 1 by sigma min. So, you get this and that brings us to the revised definition for norm and condition number of a matrix.

(Refer Slide Time: 35:40)



(Refer Slide Time: 35:44)

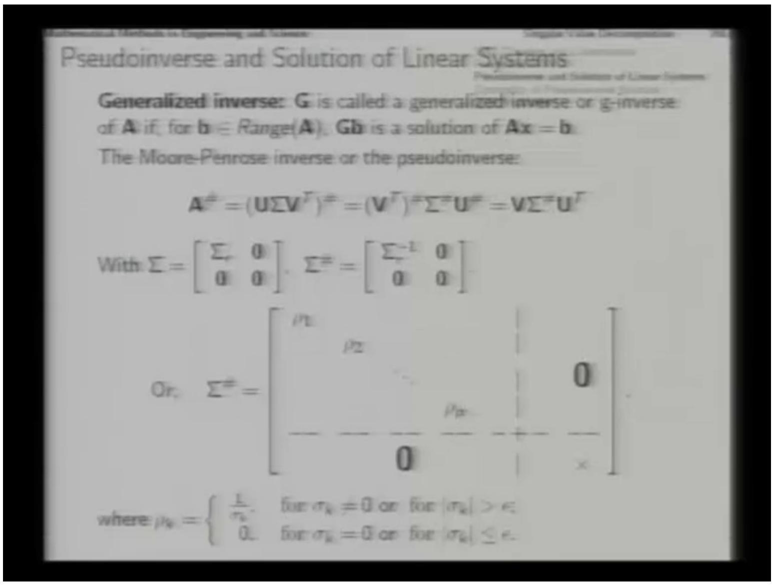


The new revised definition of norm and condition number will be like this the norm of a matrix is the largest singular value and the condition number is the ratio of the largest singular value through the least now note that this revised definition of condition number can equally cater to rectangular matrices also the old definition based on inverse would not be able to do that now note one more important issue if you can arrange the singular value increasing order as we have been talking about then with rank of the matrix has r you can write it in this manner.

So, in which U r is that sub matrix which has all the columns of U which are corresponding to Nonzero singular value similarly v r are the corresponding columns of V and U bar and V bar constitute the rest of the columns in that case this matrix A which is U sigma V transpose can be multiplied in this block form in which the 3 components that you get out of it will be 0 base because of these and the Nonzero component is only this U sigma U r sigma r v r transpose. So, the other components are 0 and this gives you this summation that will mean that if you can store the components of U and V the columns of U and V which are corresponding to Nonzero sigma then that alone will be the sigma values of Nonzero sigma; sigma k s will be able to reconstruct the matrix A and; that means, that for a large matrix which only a few top singular values as nonzero and significant you can effect a very efficient storage and reconstruction. So, with this background now go ahead and see what is the application and what is the particular advantage.

Of singular value decomposition for solving linear system of equations A x equal to b and we again revise the definition of pseudo inverse compared to what we did earlier in the chapter 7.

(Refer Slide Time: 38:15)



So, in the background there is this term called generalized inverse for any matrix you can define a generalized inverse or G inverse if for a vector be in the range a G b is a solution of this that is for a matrix A; A matrix G can be considered a an inverse of some sort generalized inverse if for a consistent right side vector b; G b gives you the solution that way G operates something like an inverse. So, pseudo inverse is actually a special case of generalized universe.

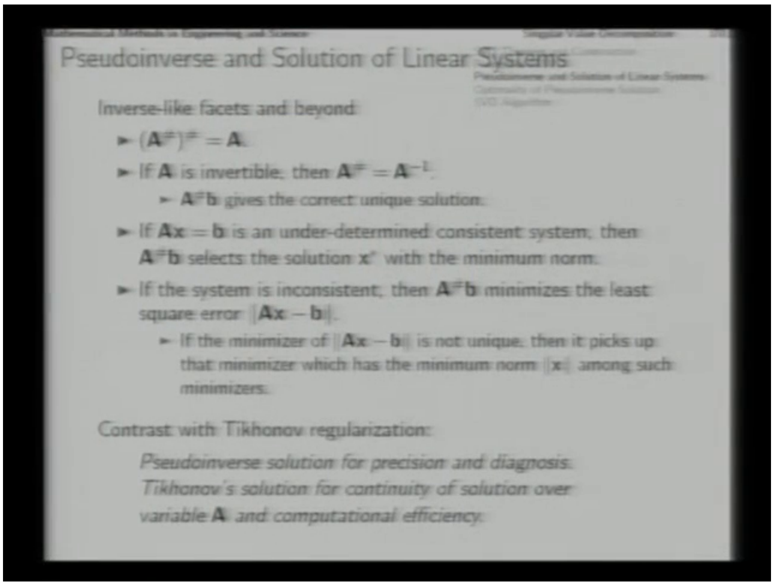
The pseudo inverse or the Moore Penrose inverse is defined in this manner and in order to differentiate it from the ordinary inverse we write it with this symbol a hash. So, a hash is U sigma V transpose hash now here when ever an inverse is actually possible we take the a hash we take the pseudo inverse as same as the actual inverse. So, the pseudo inverse of this will be V transpose hash sigma hash U hash now V transpose and U are orthogonal.

So, for them actual inverse is exists. So, for V transpose hash we write V transpose inverse which is V and similarly for U hash which is U inverse which is U transpose that actual problem which like actual problem is with this right. So, this is the one which requires a definition. So, that is defined like this for this structure of sigma in which there is a diagonal matrix of r by r size here with r Nonzero singular values and everything else is 0 sigma hash is defined as this.

So, now that will mean that those diagonal entries which are Nonzero their reciprocals will come here and those diagonal entries which are 0s. So, their reciprocal rather than infinity we put 0 here this is very interesting in place of 1 by 0 which should come as 1 by 0 by the ordinary rule here we are actually writing 0. So, this is how we define the pseudo inverse or Moore Penrose inverse in elaboration you can write sigma hash in this manner. So, sigma 1 to. So, in place of the diagonal entries row 1 to row p, you write where row k is the reciprocal of sigma k when sigma k is Nonzero and sometimes in practical cases even if sigma k is very small then we consider it as good as 0 that is here.

So, for those cases where sigma k is 0 or extremely small we put row k as 0 rather than putting 1 by extremely small number or 1 by 0 we actually put it 0 there. So, this is the definition of pseudo inverse. Now sometime at leisure you should compare this expression and this a description of the pseudo inverse with the special cases full rank cases which we worked out in chapter 7 as right inverse and left inverse. So, in those cases where the matrix is full ranked those definitions will appear as special cases of this.

(Refer Slide Time: 41:54)



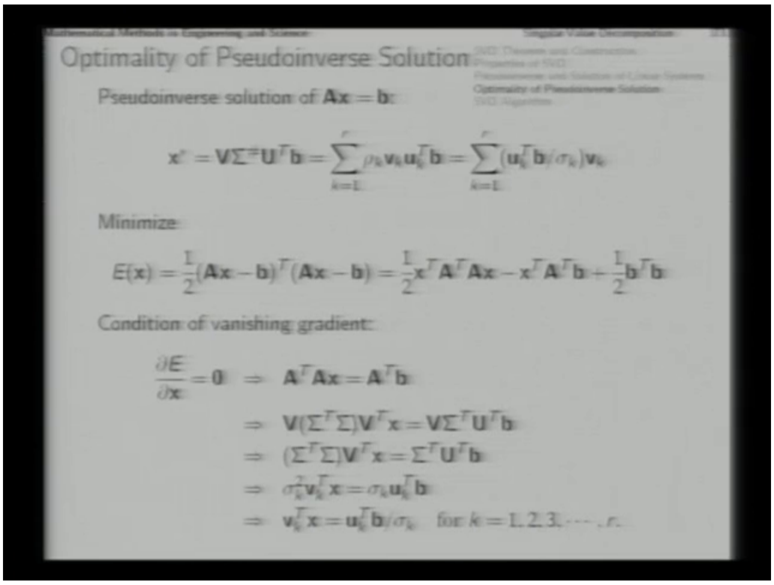
Now, what are the inverse like properties or qualities of this pseudo inverse first is pseudo inverse pseudo inverse of the matrix id the original matrix considering only actual 0 cases being put 0 here and not he truncations second important point which is like inverse that if a is actually invertible, if it is a square non singular matrix, then this will boiled on to the ordinary inverse and A hash b will give the correct unique solution of ax equal to b on the other hand if the situation is not. So, good and if A x equal to b is an under determined, but consistent system that is full rank case of more unknowns and less equations, then A hash b selects that solution x star which has the minimum norm out of an infinite possible solutions.

On the other hand if the system is inconsistent then this A hash b defined with the same formula then this A hash b will minimize the least square error that is if the system is inconsistent there is bound to be some error a x equal to b in ax equal to b ax will never be exactly equal to b, then this same A hash b will find you an x star which gives the minimum error now if that minimum error giving solution is also not unique if there are invite of them then at the same time, it will give you that solution out of those infinite possible solutions giving the minimum error which has the least size. So, all these sensible things the pseudo inverse does with the help of a single definition now you should contrast this with the solution which is obtained earlier from Tikhonov regularization.

So, pseudo inverse solution is typically used when you want precise values and also for diagnosing a linear system whether it has any such inconsistency or under determinacy problems and s, on; on the other hand, Tikhonov solutions can be used when the position matrix A changes over a domain and you want continuity of solutions.

So, Tikhonov stations is preferable for continuity, but diagonals is and for precise solutions pseudo inverse solution is better Tikhonov solution will always inhibit some error. Now in he exercises of this chapter in the text book actually the reason its exercise which asks you to determine the Tikhonov solution and the pseudo inverse solution and compare then for a matrix A which has one of the components variable now we want to know how this whole thing is accomplished by a single formula.

(Refer Slide Time: 45:12)



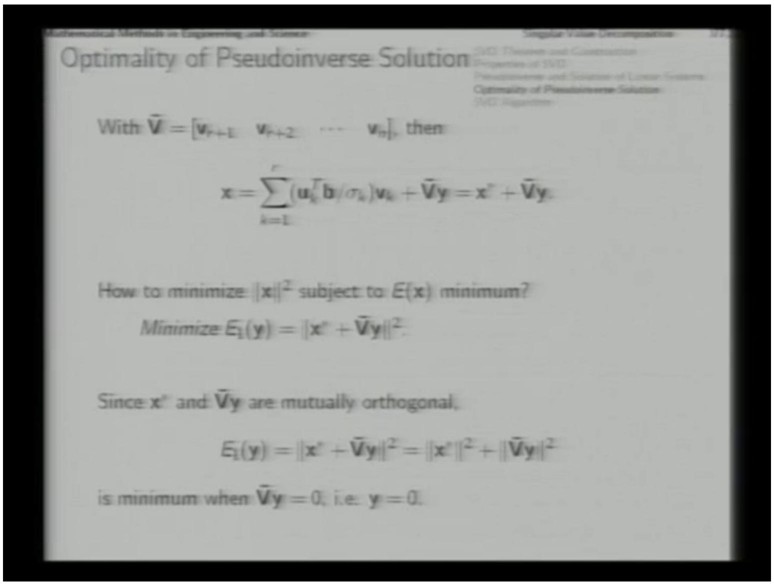
So, for that first, we note down what is the pseudo inverse solution that we find that is this is the pseudo inverse of A and when we multiply it with b we got this sum where the summation is over k from one to r that is for all the Nonzero singular values. So, for that we get this expression and when we reduce it, then we have U k transpose b which is scalar divided by sigma k because row k is 1 by sigma k.

We can if we write it like this then we will find that the pseudo inverse solution that we are getting is actually a linear combination of r basis members v 1 to b v r the corresponding components of these scalar values written in the parenthesis now we want to pose the problem as first minimization of the error and then if the solution is infinite then further minimization of the size if the solution and then see whether we get this same solution. So, if we want to minimize the error these square error half norm square of the error a x minus V, then as we open this we have already encountered this earlier ones then the minimalist condition first order condition is that its derivative its gradient with respect to x must be 0. So, when we do that we get this as we got last time here now in place of a we write U sigma V transpose and through few steps we come to this point now note that this is a matrix equation.

And this is the corresponding scalar equation for each component of that vector equation right. So, this is for each k from k equal to 1 to r where r is the rank that is Nonzero singular values. So, from here that you find that V k transpose x that is component of x along the unit vector V k turns out to be U k transpose b divided by sigma k this sigma k square goes down in the denominator and this is what is actually sitting here right. So, in this solution x star is composed of several vectors v 1 v 2 up to v r in which the component of V k is this right.

So that means, x star is actually giving you this combination of these vectors with these components now this first order condition for the minimality of this tells you what should be the components of the solution along the basis vectors v 1, v 2, v 3 up to v r along v r plus 1, v r plus 2, v r plus 3; what should be the component that is not mentioned here; that means, those components can be anything the error is still remaining because the condition is satisfied; that means, the general solution for minimum error you can constitute with the components along v 1 V to v r as specified here and any component along the rest of the directions that will give you this with components as prescribed along the first R directions along the first R basis numbers and anything in the rest; that means, y is free here. So, v r plus 1, y 1 v r plus 1 R plus 2 y 2 and so on.

(Refer Slide Time: 48:43)

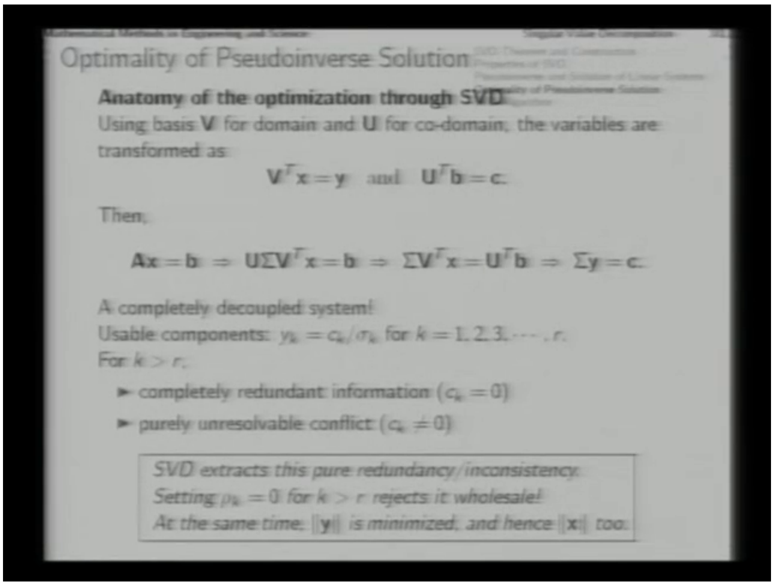


Since these y 1, y 2, y 3 can be anything the V bar is the basis for the null space that you will appreciate because any null space member will not change anything in the solution in the right hand side. So, now, we say that out of all these infinite possible solutions which one is the one which is of least size. So, then what we ask for we ask for how to minimize the size of the vector subject to this error being minimum anyway; that means, the solution you take from here and minimize it with respect to y that is which y to select to minimize the size of the vector this. So, we say minimize the size that is x norms square that is this.

So, now you find that x star this part is a linear combination of v 1 to v r and this part is a linear combination of other basis members and all other basis members are orthogonal to the basis members of the first family that will make this x star sitting in one sub space and this part V bar y sitting in another sub space 2 sub spaces being orthogonal to each other. So, how do you find the square of some of this if the 2 members are in a in orthogonal sub spaces. So, since they are mutually orthogonal this will be simply x star square plus V bar y norm square now you find that if we then want to ask that which y will give this as minimum where this is already available and cannot be tempered only y can be changed then y equal to 0 will give you this as 0 and this sum as minimum.

So; that means, that y equal to 0 will you give you the minimum size vector x which is of this form which minimizes the error.

(Refer Slide Time: 51:13)



Now, how this whole thing happens that you get all the optimal conditions in the suggestion that you construct with the help of the pseudo inverse. So, for that let us investigate the anatomy of this optimization through SVD if we use basis V and basis U for the domain and co-domain then the variables x and b under question x unknown b right hand side unknown they are transformed as this that is in the new basis V the expression of x will be this y and in the new basis U for the co-domain the vector b will be represented as c which is U transpose v.

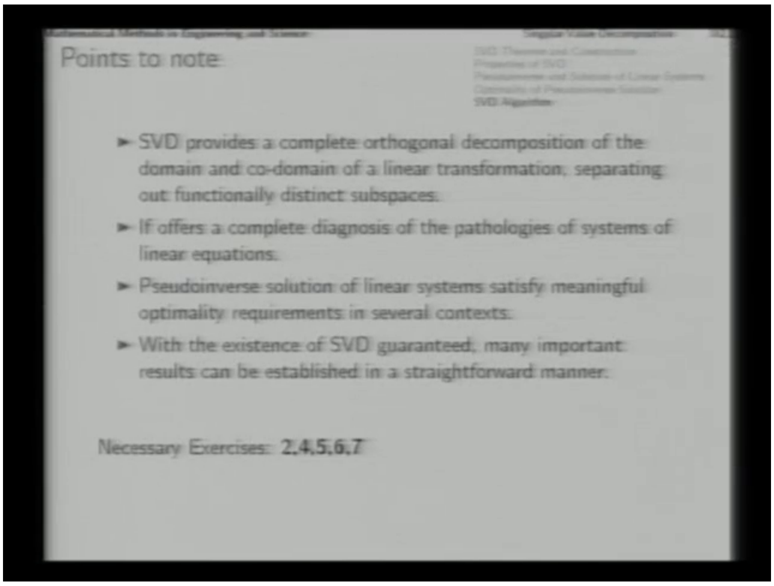
Now, if we write the system of linear equation a x equal to b and a is U sigma V transpose then V transpose x is y and U brought here as U transpose multiplied with b U transpose b is c. So, then you basically get the equation in the new bases V on this side and U on this side as sigma y coequal to sigma c and this is a completely decoupled system because if we write this system of equations sigma y is equal to c we will find sigma 1 sigma 2 up to sigma R like this y 1 y 2 up to y r and below that possibly more variables up to y n and on this side we will have c one c 2 V up to c R and below perhaps more things now the way the singular value solution has been constructed you get the useful information only from the first R rows first R equations and they are completely decoupled because y 1 simply c 1 by sigma 1 y 2 is simply c 2 by sigma 2 up to this y i is c R by sigma r.

What happens below; you find that all 0s here; that means, right left side of the equation will give 0 question is what is here in c if there are corresponding 0s here then that will means that the system is consistent, but that information 0 equal to 0 is completely unusable if it does not have any information content on the other hand if some particular values here R Nonzero that will mean we are talking about 0 equal to something Nonzero; that means, that is the conflict that is the source of inconsistency in the system of equations.

So, in this situation we find that for k equal to one to R this is what we determined and that is the only useable component and for c k greater than 0 for k larger greater than one R that is below for c k greater than Nonzero for c k is Nonzero you will find that you have purely undesirable conflict s that is simply the inconsistency decompose into an orthogonal sub space and which cannot be compensated for by any other component and c k equal to 0 will give you completely redundant information that is again the completely redundant information is also collected over an orthogonal sub space which cannot be changed from any other component from outside.

So, by setting the appropriate diagonal entries of sigma hash as 0 SVD extracts this pure redundancy and inconsistency and rejects that. So, it rejects the redundancy it rejects the inconsistency and gives you that solution which is the best possible achievable at the same time since these were free skill because the usable component gave you the value of only this much setting this variables as 0 minimizes the norm of y and since the norm of x will be the norm of V y V is orthogonal. So, though the multiplication of an orthogonal matrix the norm of the vector does not change. So, minimum y will mean minimum x for the norm.

(Refer Slide Time: 56:09)



Now, the points to notice here important points to note are the following the first SVD provides you a complete orthogonal decomposition of the domain and co-domain and it separates functionally distinct subspaces on this side the null space from the rest on that side the range from the rest it offers a complete diagnosis of the pathologies of a system of linear equations and then pseudo inverse solution A hash b gives you a the most meaningful solution of a linear system in all cases apart from these what has not been noticed still now clearly is that with the existence of SVD guaranteed that any metrics real or complex you can write as U sigma V star or V sigma V transpose many important mathematical results and many other formulations can be worked out in a straightforward and direct manner in many of the cases in coming lectures based on this existence of SVD you will find that you will be able to appreciate the deductions of many of the results quite easily.

So, here we in this lecture we have actually connected 2 important problems systems of linear equations and Eigenvalue problems together through the singular value decomposition in the next lecture which will be the last lecture of our linear algebra module we consolidate a few important issues based on the abstract fundamental ideas of linear transformations.

Thank you.