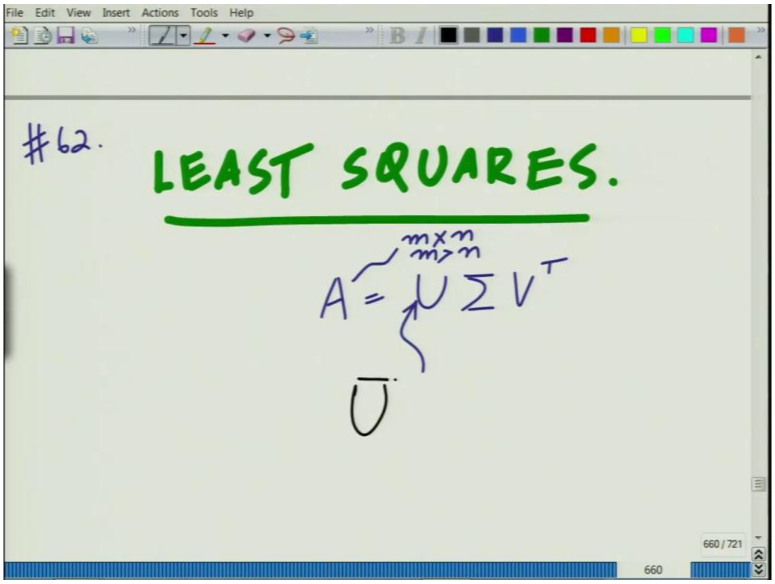
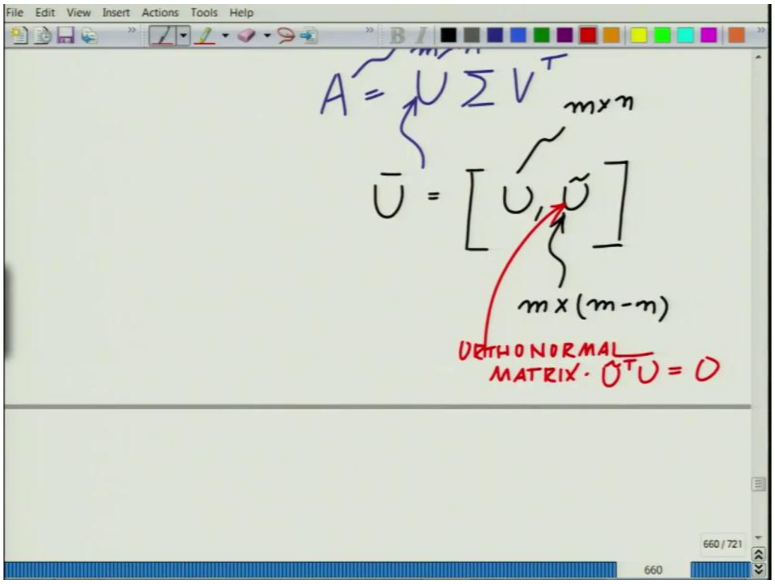
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Lecture 62
Least Squares using SVD

Hello, welcome to another module in this massive open online course. So are we looking again, putting a relook at the least squares, especially for the scenarios where the matrix A transpose A is not invertible that is A does not have full column rank.

(Refer Slide Time: 00:31)

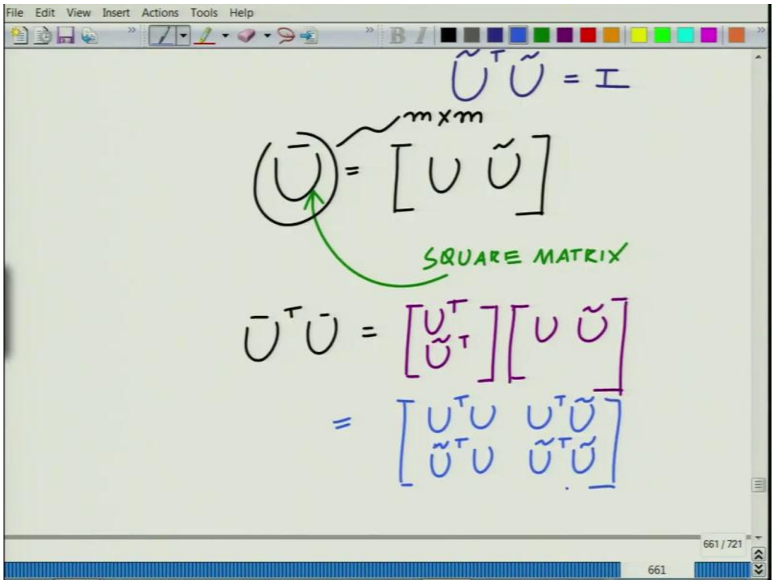


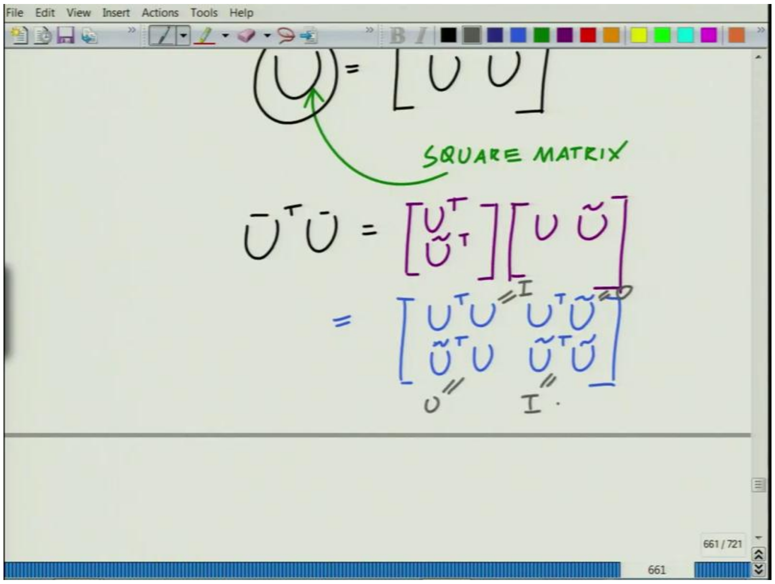


So, let us look, so we are revisiting our least squares problem. And, so we have A equal to U sigma V transpose A is a tall matrix, A is m cross n with m greater than n. And, and therefore, I can expand U as follows that is what we are saying U bar which is basically equal to the matrix U, U tilde.

So, this is your m cross n matrix and this is your m cross m minus n matrix. A And I can always find, so U is also an orthonormal matrix. So, U tilde is also an orthonormal matrix such that U tilde transpose U equal to 0 that is U tilde as columns essentially that are orthogonal to U. So, U tilde transpose U equals 0.

(Refer Slide Time: 02:18)

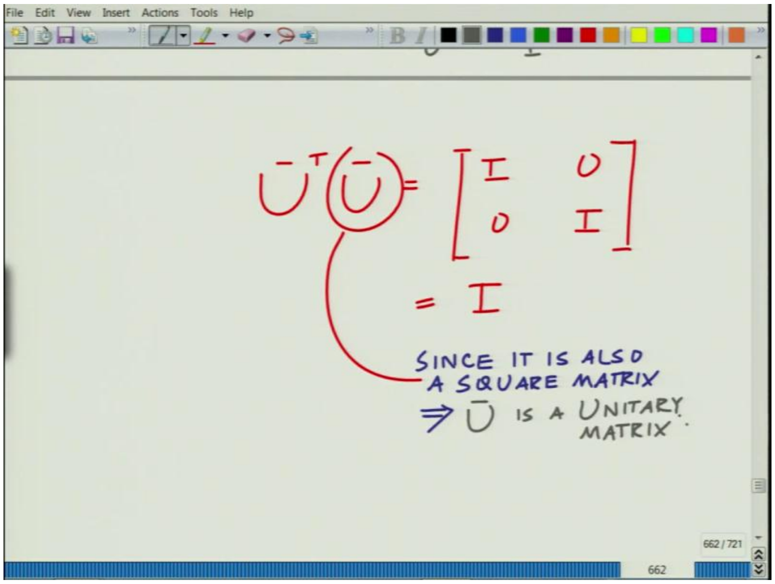


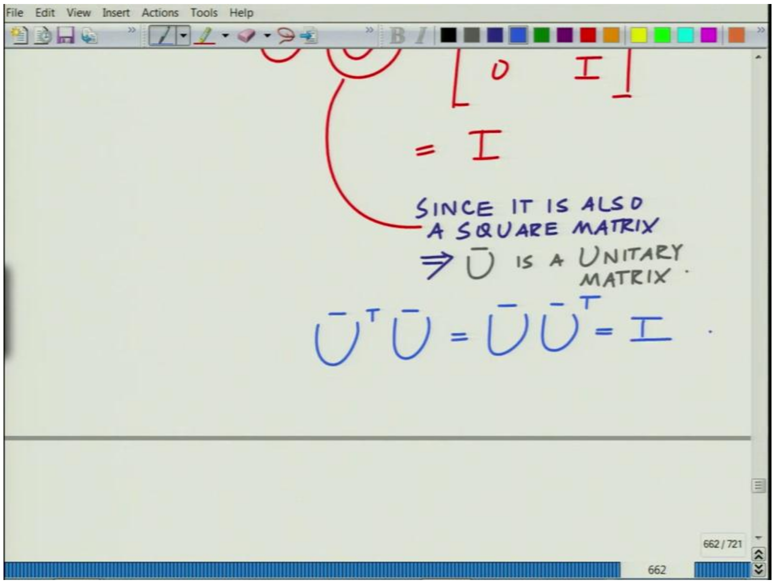


And further U tilde, because it is orthonormal automatically implies that U tilde transpose U tilde equals identity. So, now, we have this matrix U bar which is essentially if you look at this, this is U U tilde which is this is therefore, an m cross m matrix or this is essentially a square matrix. So, U U bar is essentially, this is essentially a square matrix.

So, this is essentially a square matrix and if U look at now, U bar it satisfies the property it is not very difficult to see that U bar transpose, this will be equal to U transpose U tilde transpose U times U tilde. And which now if U simplify this interestingly this will be, well U transpose U which will be identity U transpose U tilde which will be 0 U tilde U tilde transpose U and U tilde transpose U tilde which will again be identity. So, if U look at this matrix, so in this matrix this is equal to identity, this is equal to 0, this is equal to 0 and this is equal to identity.

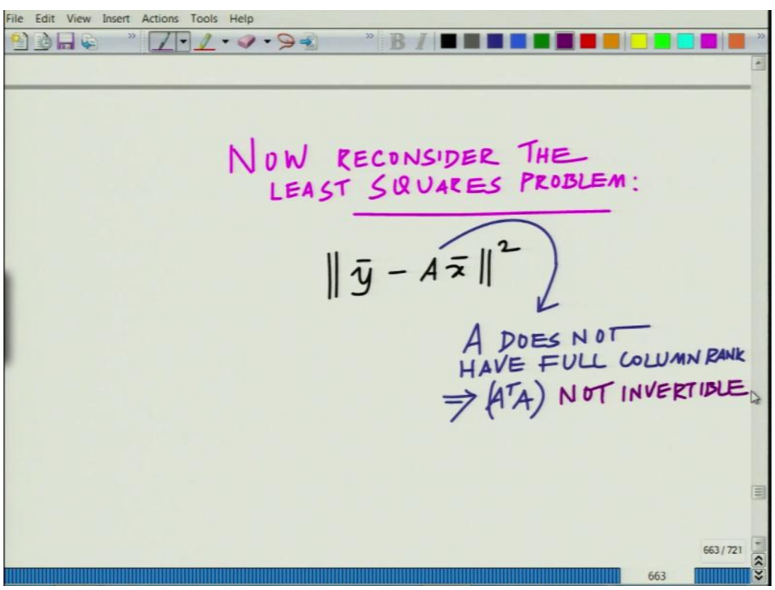
(Refer Slide Time: 04:18)

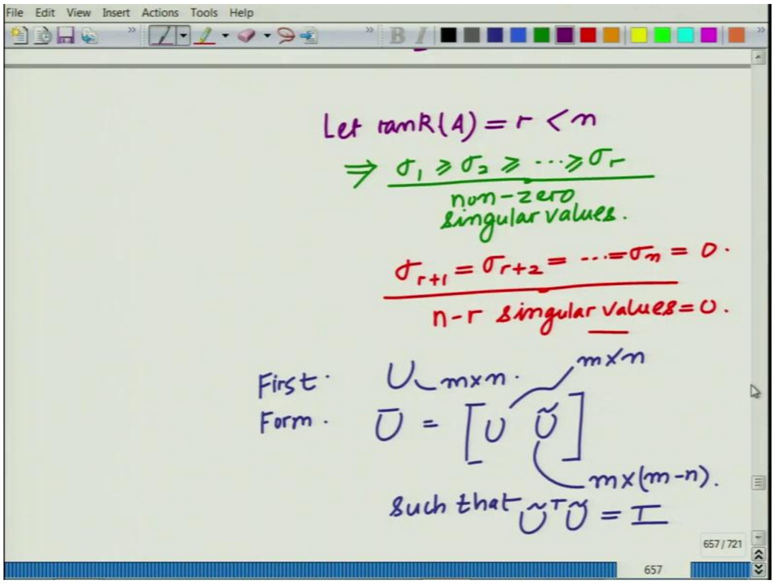


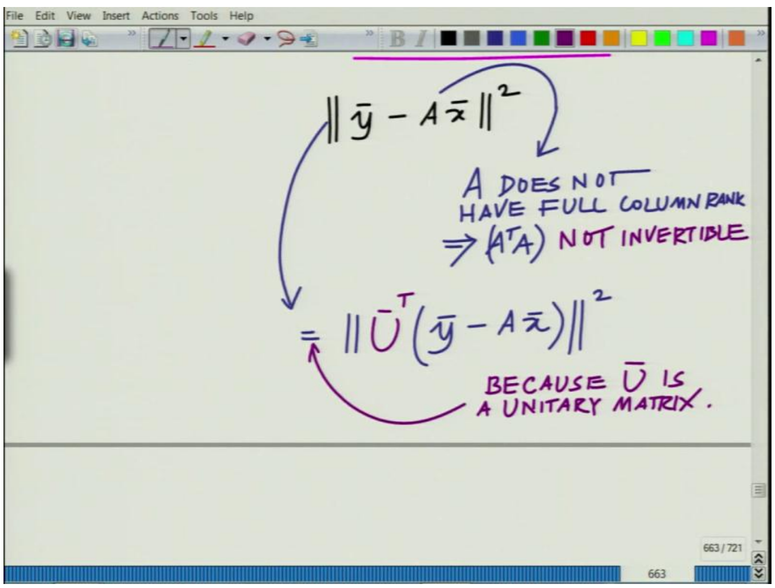


So, if I write it out, you will have U bar transpose U bar equals identity 0 0 identity which is essentially one big identity matrix. So, essentially what it implies its square and satisfies the property U bar transpose U bar equal to identity. Since it is also square, since it is also a square matrix, this implies U bar is a unitary matrix, which means that U bar transpose U bar equal to U bar U bar transpose and both of these quantities are equal to identity. So, we will have U bar transpose U bar equals U bar U bar transpose and both these quantities are equal to identity.

(Refer Slide Time: 05:43)





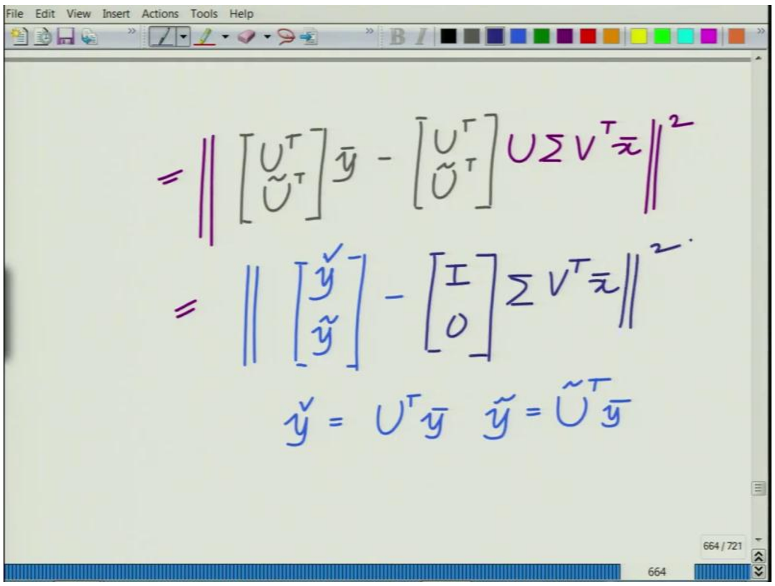


And now, let us consider, let us go back to our least squares problem. Now, I would say reconsider our least squares, reconsider the, now reconsider the least squares problem that is we have remember norm y bar minus our Ax bar square, where A now we are considering the special case where A does not have full column rank, this is the A does not have full column rank implies A transpose A is not invertible.

And that is the reason we cannot do the usual A transpose A inverse A transpose y bar. So, our A transpose A is not invertible. In fact, we had seen this if U look at the singular value decomposition we have n minus r singular values that are 0. So, r non zero singular values and n minus r singular values which are 0. And therefore, now if you go back and take a look at it what is going to happen is essentially you will have this can be given as norm of U transpose.

Now, I can simplify this least squares norm y bar minus x bar square, I can write this as interestingly, I can write this as is equal to norm U transpose and this is key bar minus A x bar whole square y, y is the norm of this vector equal to norm of this vector because U bar, U bar is a rotation matrix because U bar is a unitary matrix this norms, both these norms are equal because, why are these equal, because equality follows because U bar is a unitary matrix.

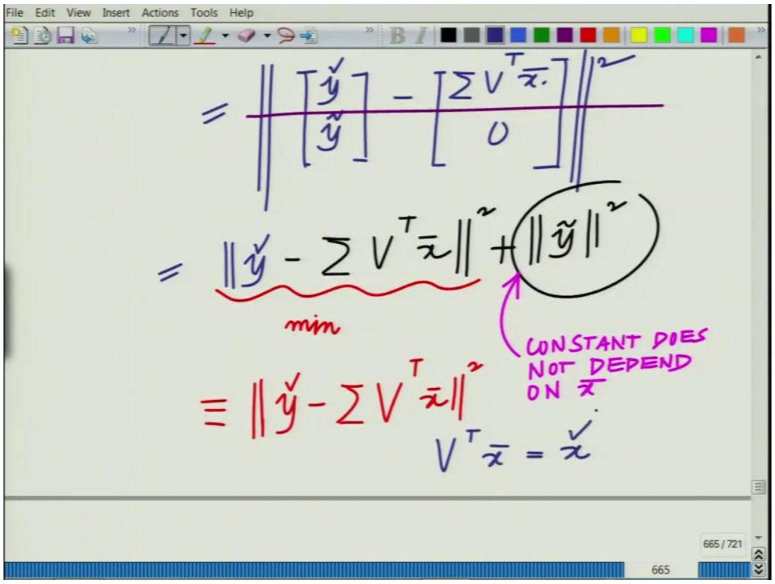
(Refer Slide Time: 09:01)

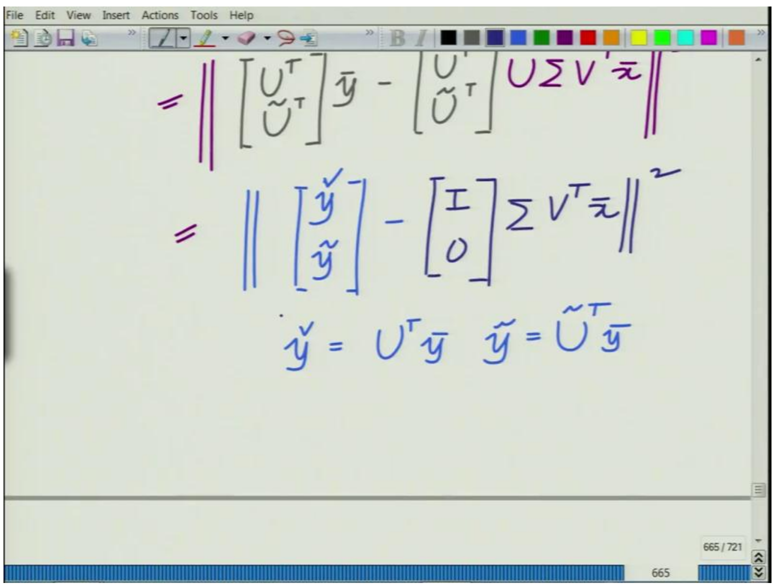


So, this implies, now if you look at this, this implies, this is norm of U bar transpose let us decompose y as, so now U bar transpose is nothing but, if you write it in terms of its components this will be remember, U transpose U tilde U tilde transpose y bar minus A again, U transpose, sorry, again U transpose U bar transposes U transpose U tilde transpose times U sigma V transpose into x bar norm square, that is essentially what you have.

And now if you multiply these two quantities, now you will see something interesting. So, you will have the norm of, let U transpose y bar, we can call this as y check and U tilde y bar you can call this as y tilde. So, y check equals u transpose y bar y tilde equals U tilde transpose U tilde transpose y bar minus you will have, this is interesting U transpose U this is equal to identity. And U tilde transpose U this equal to 0. So, identity into, identity 0 times sigma V transpose norm x bar square.

(Refer Slide Time: 11:16)





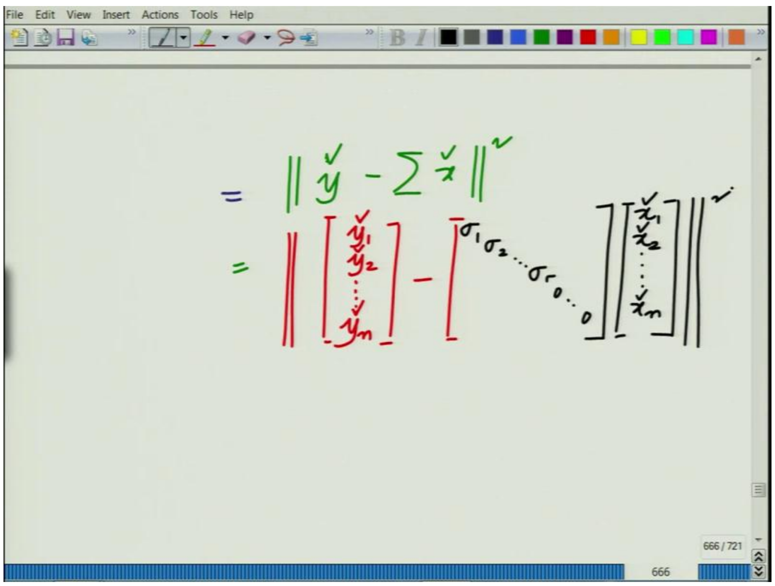
Which now you can simplify it, which is now you can see you can write it as two things, which is norm of if you look at the top part, or let me just write one more step. So, you have y check y tilde minus this our identity into sigma V transpose x bar. So, that will be sigma V transpose x bar 0 times square. And now, we can partition it into two parts.

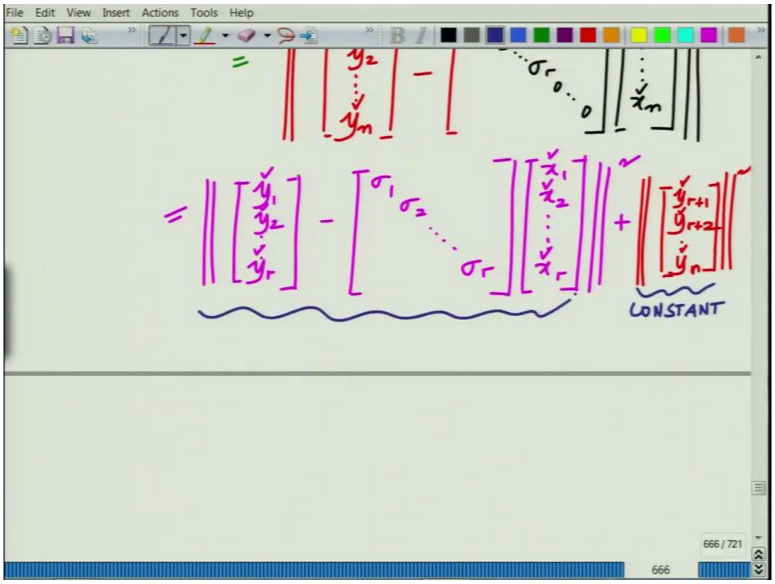
So, this is the norm square. So now, if you partition this into two parts, you can write this as the norm square of, so we take the norm square of a vector that is the norm squared of the top part plus the norm square of the later part. So, this will be essentially, this is essentially the norm square of y check. I would say minus sigma V transpose x bar square plus norm check, norm of y tilde square.

Now, this is a constant, now you can see this is constant does not depend on x bar which means we are left because this, remember we are trying to find the x, x bar which minimizes the least squares cost function, but this right part that is fixed that cannot be minimized, that does not depend on x bar. So, we are left with the left, that is your norm y check minus sigma V transpose x bar square.

So, we are left with this. So, I can write this as the equivalent problem remember this three equal to signs, three horizontal bars this denotes the equal to sigma V transpose x bar square. Now, we set this V transpose x bar equal to x check, the vector x check.

(Refer Slide Time: 13:34)

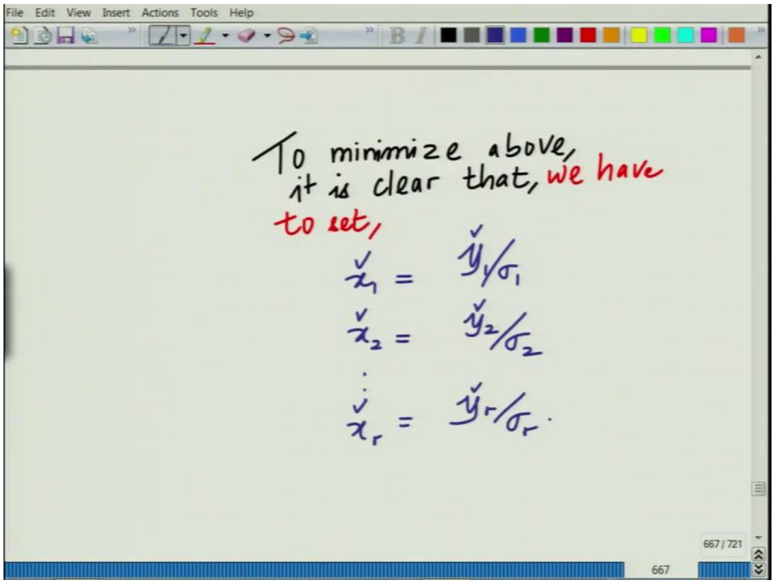


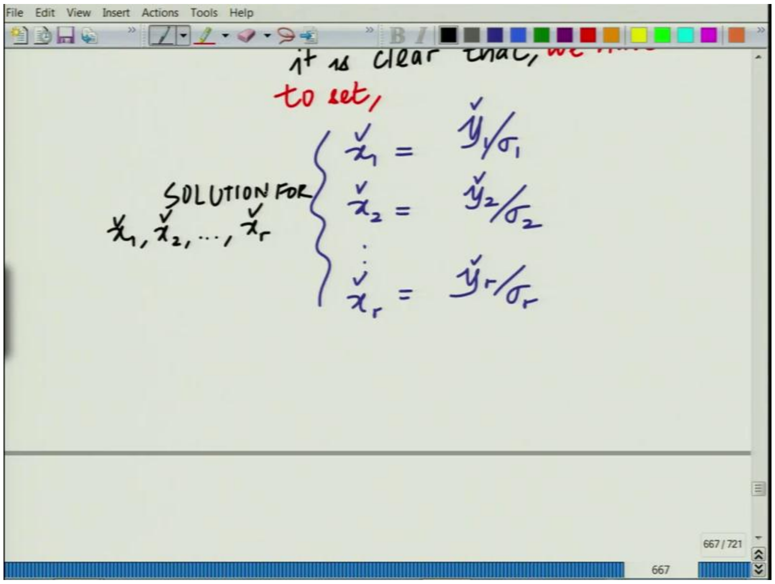


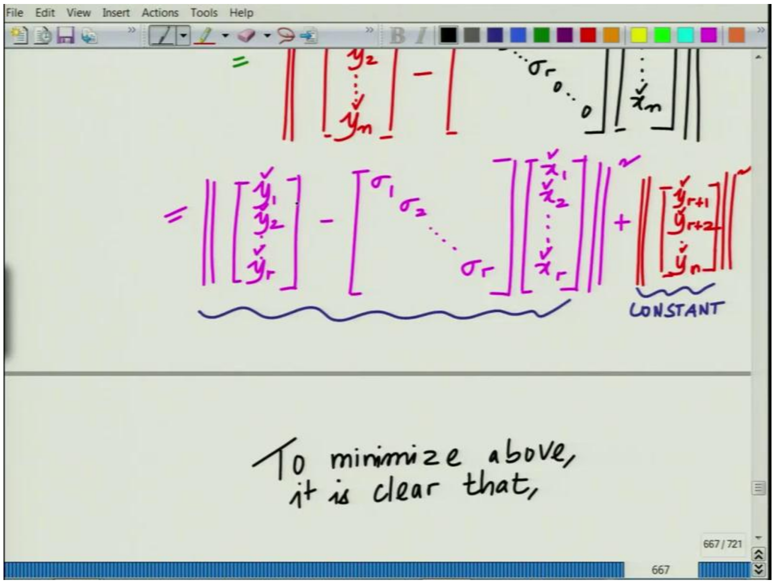
And this becomes, if you look at this, this becomes your y check minus sigma x check square which is essentially if you look at this, you will have this one, you will have y1 check, y2 check, so on up to yn check minus the sigma will be, remember sigma 1 the non-zero singular value, sigma 1, sigma 2 up to sigma r followed by n minus r 0s. And then you will have the x 1 check x 2 check, so on to xn check.

Which I can now write again as two parts. Now, if you look at this I can again write this as two parts. So, that will be the first part, the top part corresponding to the non-zero singular values y1 check, y2 check, so on, yr check minus sigma 1, sigma 2 to sigma 1 r x 1 check, x2 check up to xr check square plus this will be this vector which is your yr or xr plus 1 or this will be your, further this will be your y r plus 1 y check r plus 1, y check r plus 2, so on, y check n whole square. And once again this is a constant, does not depend on x check.

(Refer Slide Time: 16:48)



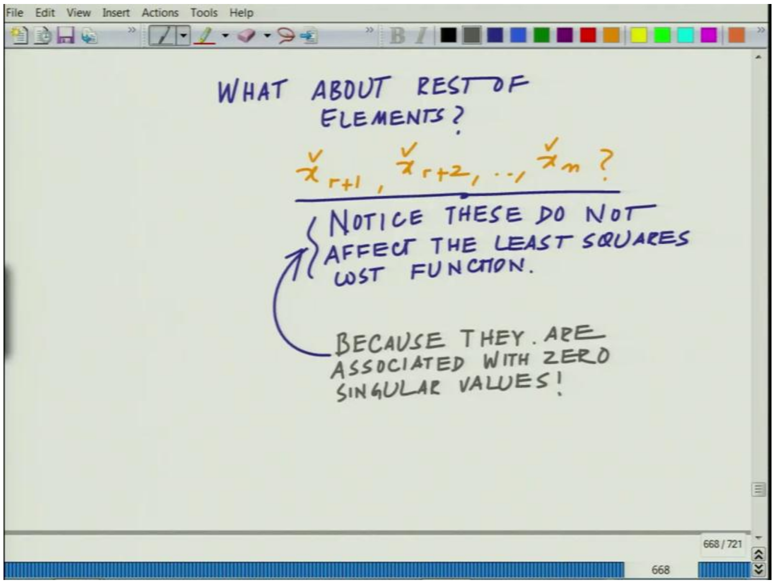


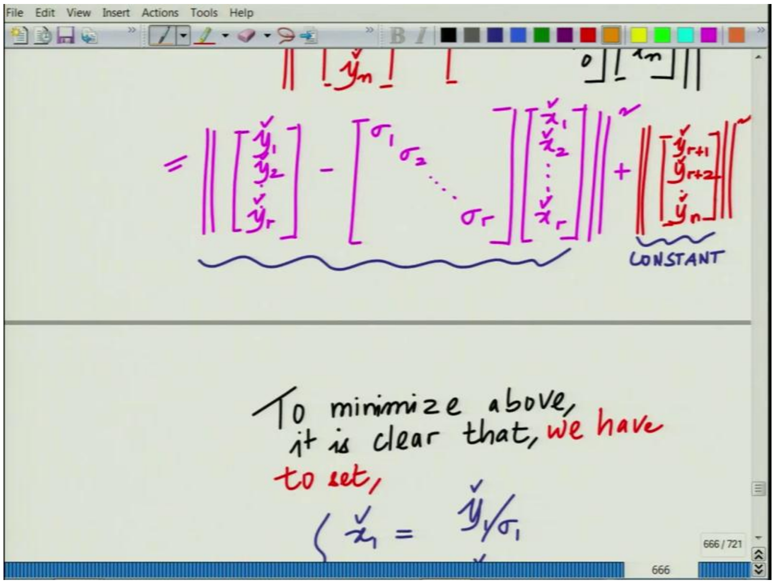


And therefore, now to minimize this, the only thing that I can do now to minimize this, now to minimize the above it is clear that I simply can make this 0. How can I make this 0, I simply set y1 check equal to sigma 1 to x1 check, so I can check x1 check equal to y1 check divided by sigma 1. Similarly, x2 check can be made equal to y2 check divided by sigma 2. So, that is essentially how to do it.

So, it is clear that to minimize above, it is clear that we have to set, what do we have to set, it is very clear that we have to set x1 check equal to y1 check or sigma 1 x2 check equal to, so on, xr check equal to yr check divided by sigma r. So, these are the solutions, this is the solution for the x1 check, x2 check. So, this is the solution for, now, that begs the question what about the rest of the elements, that is what about the rest of the elements that is xr plus 1 check, xr plus 2 check, and xn check.

(Refer Slide Time: 19:04)

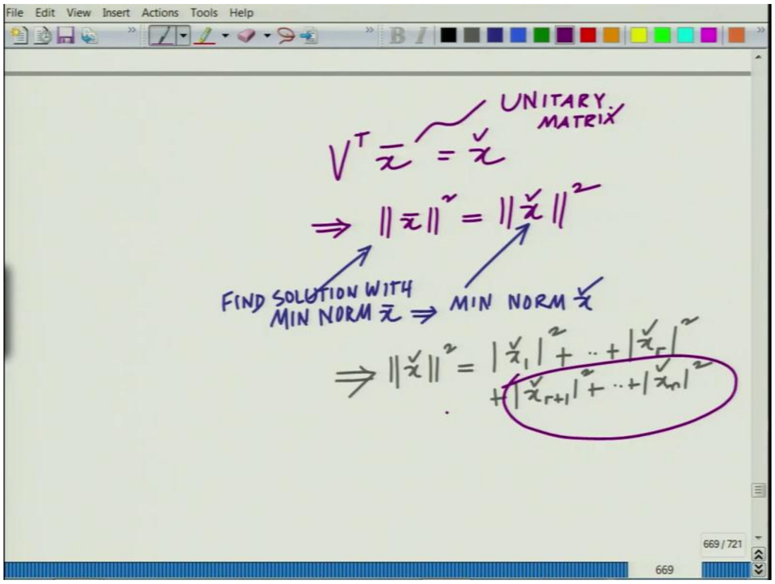


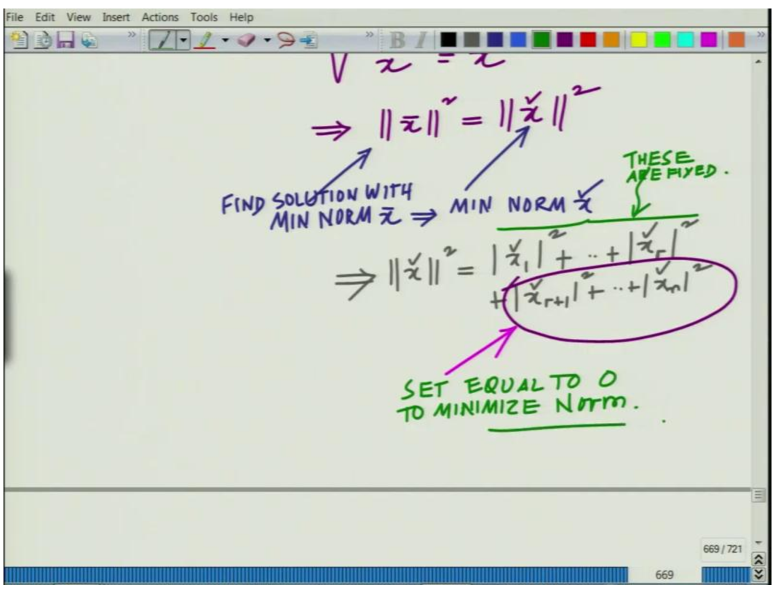


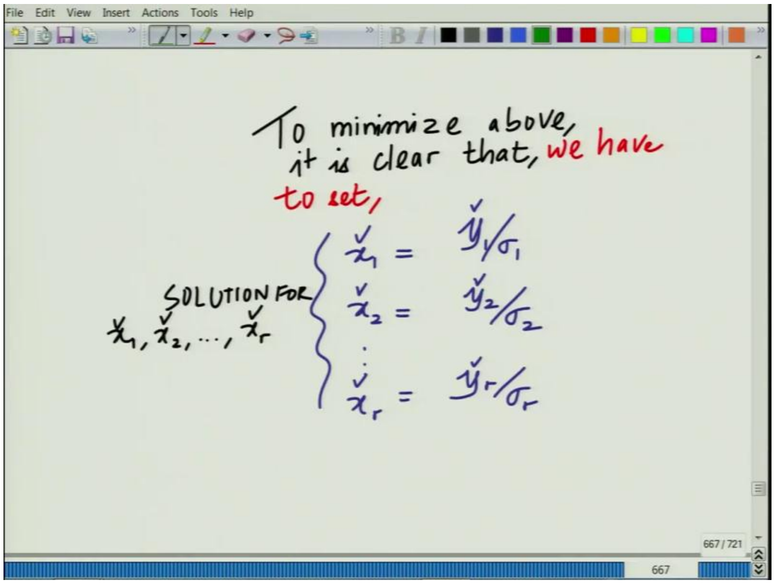
Now, what about, what do you mean by rest of the elements that is we have the x r plus 1 check, x r plus 2 check, so on up to xn check. Now, what about these, now notice that these do not affect the least square solution, because it does not depend on this. So, notice that you can see from here, this does not affect the least squares that is, it does not affect the error.

Notice these do not affect the least squares cost function, why, because they are associated with zero singular values, because they are associated with zero singular value. So, now you can set them to anything, that is the truth of it. So, these can be set of it anything. How do we set them in practice, we set them such that norm of, now therefore, we can find, (infinite), now this means that we can find an infinite number of solutions by setting this x check r plus 1, x check plus r 2, x check n arbitrarily which gives the same least squares error.

(Refer Slide Time: 21:40)







So, how do we set them, the point is not to understand that now, we go back and take a look at it out of all these solutions which one is the solution that we prefer, we have V t times x bar equal to x check. Now look at this, how do VT times x bar equals x check. Now, remember this is a unitary matrix and this implies norm of x bar square equal to norm of x check square, because once again multiplying by unitary matrix or simply rotation matrix complex rotation matrix or any rotation matrix does not affect the norm.

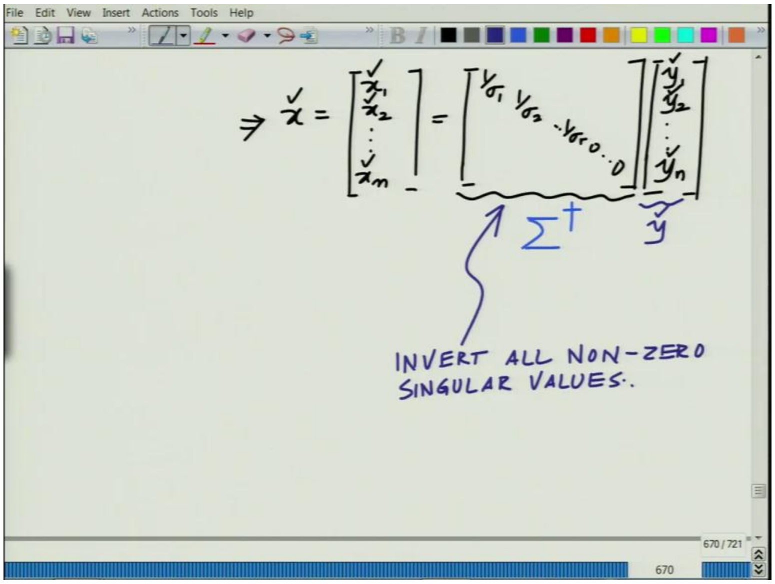
And therefore, now how to check, how to set it. We want to set minimum, minimum norm of x bar, we want to demonstrate or we want to get minimum norm of x bar which implies because these norms are equal, we want to find the solution which has the minimum norm of x check. Solution, with, so we want to find the solution with minimum norm x bar which means the minimum norm of x check which implies.

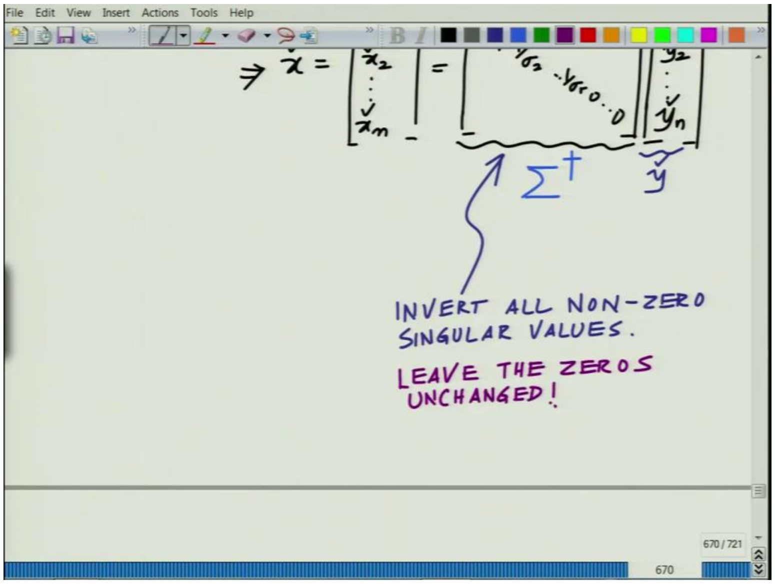
Now, your norm of, remember your norm of x check square equals norm of x check 1 or magnitude, rather magnitude x check 1 square, so on until magnitude x check r whole square plus magnitude x check r plus 1 whole square until magnitude x check n whole square. Now, since these do not affect, since these do not affect the cost function set these, to minimize the norm set these to 0, that is, that is a simple.

So, these are not going to affect the cost function anyway. So, they are unnecessarily contributing in the norm, to minimize the norm set these equal to 0, since these do not otherwise affect the cost, these can be set arbitrarily remember the cost remains the same. So, set them equal to 0 to minimize the norm. So set these equal to 0.

Now, if you look at the 1 x check, 2 x check r these are fixed, these we cannot change because remember x check 1, x check 2, x check r these are fixed, that will be given by y check, 1 by sigma y check, 2 by 2 sigma y check r by sigma, so these cannot be fixed. So these are fixed. The variable one, the arbitrary ones we set to 0 to minimize norm, so that is the point. So, these are fixed. So these are fixed, these you set equal to 0 to minimize the norm.

(Refer Slide Time: 25:03)

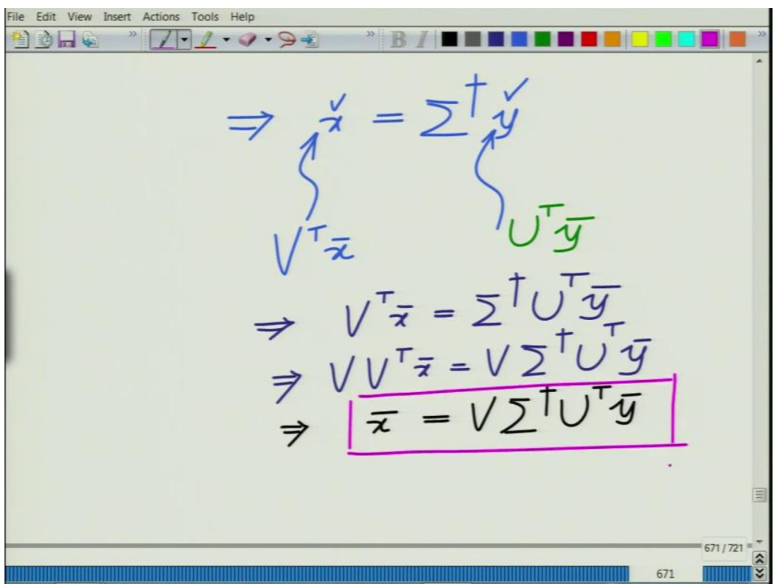


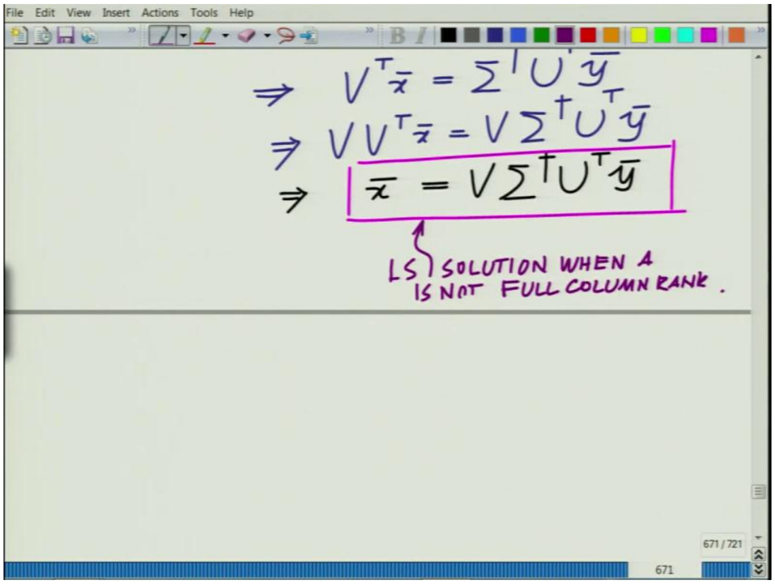


Implies we get, now finally we get, what do we get, we get x check which is the vector x1 check x2 check, so on up to xn check equals I am going to write it down as follows 1 over sigma 1 1 over sigma 2, so on, 1 over sigma r, rest will be zeros times your y1 check, y2 check, yn check. And this is a very interesting, that is what it means is all the non-zeros singular values you are inverting, all the rest of the 0 singular values we are remaining unchanged.

And this is essentially what is called is, slightly different from your sigma inverse, this is what is called as sigma dagger, the pseudo inverse of sigma, that is essentially what you are doing is and this is your y check, what is the sigma dagger, essentially invert all non-zero singular values, all the non-zero singular values you invert. Rest the 0 singular values leave the zeros unchanged, leave the zeros unchanged.

(Refer Slide Time: 27:10)



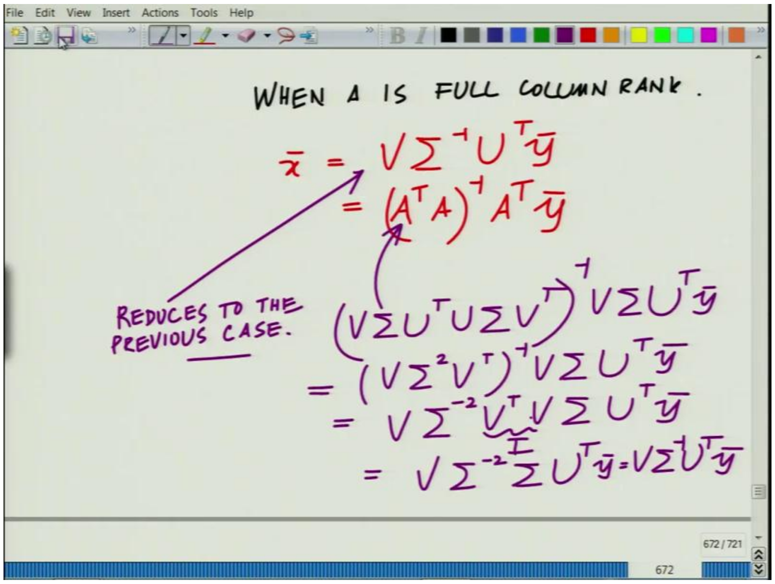


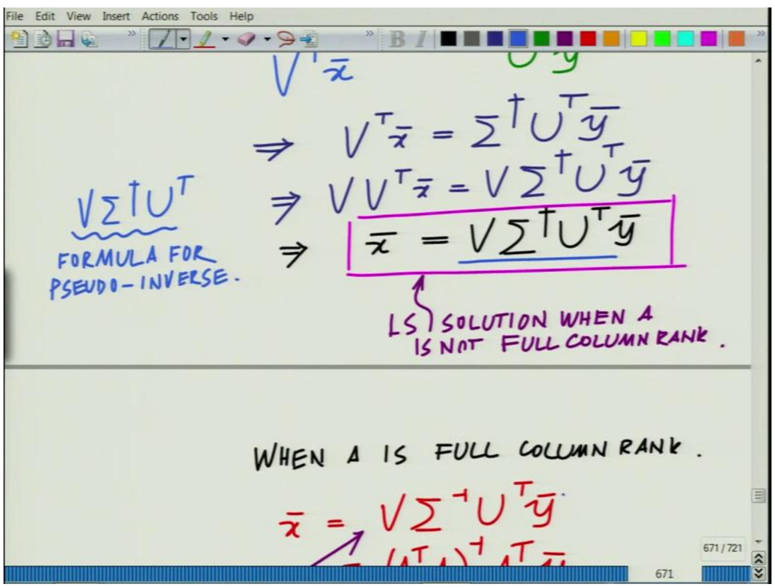
And that implies essentially that you have this nice solution where you have x check equals sigma dagger y check. Now remember, x check is V transpose x bar y check is essentially nothing but U transpose y bar. So, this is essentially you, you realize that this is your U transpose y bar. So, this implies essentially that your V transpose x bar equals sigma dagger U transpose y bar and V transpose in inverse, we multiplied by V on both sides.

So, we have V V transpose x bar equals V sigma dagger U transpose U transpose y bar which essentially implies V V transpose identities. So, x bar equal to V sigma dagger U transpose U transpose y bar or this is essentially the least squares, essentially the least square solution, so very simple this is the least square solution, LS solution when A is not full column rank.

And this also you can say the solution even when A is full column rank, because when A is full column rank, remember sigma dagger simply reduces to sigma inverse that is you, since all the singular values are non-zero, you invert all the non, all the singular values.

(Refer Slide Time: 29:14)





So, when A full column rank, and not very difficult to see when A is full column rank, you have x bar equal to V sigma dagger becomes sigma inverse U transpose y bar which is not very difficult to see that this will be A transpose A inverse A transpose y bar because if you look at A transpose A inverse substitute the singular value decomposition this will be your U sigma A transpose is A is U sigma V transpose A transpose is V sigma U transpose A is U sigma V transpose inverse V sigma U transpose A bar.

So, U transpose U is identity. So, this is equal to V sigma square V transpose inverse V sigma U transpose y bar, it is in a, V is a unitary matrix. So, this is V transpose inverse, so this is V sigma square, remember none of the singular values are 0. So, I can write the sigma to the power minus 2 V inverse which is V transpose V sigma U transpose V bar. Once again, V is unitary matrix V transpose V is identity.

So, this is V sigma minus 2 sigma U transpose y bar which is nothing but V sigma inverse U transpose y bar. So, this reduces to your, essentially reduces to the, reduces to the conventional LS solution or reduces to the what we have known before, reduces to the previous case. In this case it reduces to the previous case, that is it.

So, essentially this is a very, very, and therefore, this you can now term as the general, a pseudo inverse when or this quantity that is V sigma dagger U inverse. So, V sigma dagger U transpose this is the, now this is your general formula for the pseudo inverse which is valid even when the singular, sum of the singular values are zeros, formula for pseudo inverse. This is the general formula for the pseudo inverse, which is valid even when some of the singular values of A are 0, i.e. in other words when A is not full column rank, i.e. when A transpose A or if a is complex A Hermitian A is not inverted. That is essential the theory.

So, this is an interesting extension and I think it arises fairly frequently in practice, it is good to know such corner cases, such special cases because such matrices in such applications, such scenarios can arise very frequently in practice, because it is not guaranteed that the matrix A which we also call us the sensing or the dictionary matrix, or whatever essentially is, is not always guaranteed to have full column rank. So, let us stop this module here and continue with other aspects in the subsequent modules. Thank you very much.