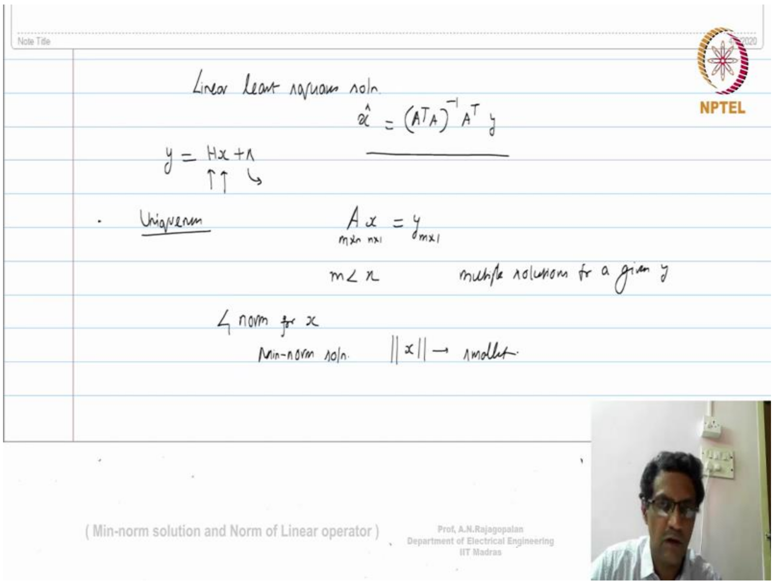
Image Signal Processing
Professor. A. N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture No. 74
Min-norm solution and Norm of Linear operator

(Refer Slide Time: 0:18)



So, we saw already what is linear least square solution or what is called LLS, linear least square solution. And which we saw turned out to be x hat is equal to A transpose A inverse A transpose y. And in order to relate it to let us say deblurring problem where you can think of A as your volume matrix where y is equal to H x plus n, and because of those noise it takes you away from this span. So, y is no longer in this span of the column of H. And that for it one would then, would not want to ask for the solution x such that the norm of y minus Hx square, norm y minus Hx square is small as possible.

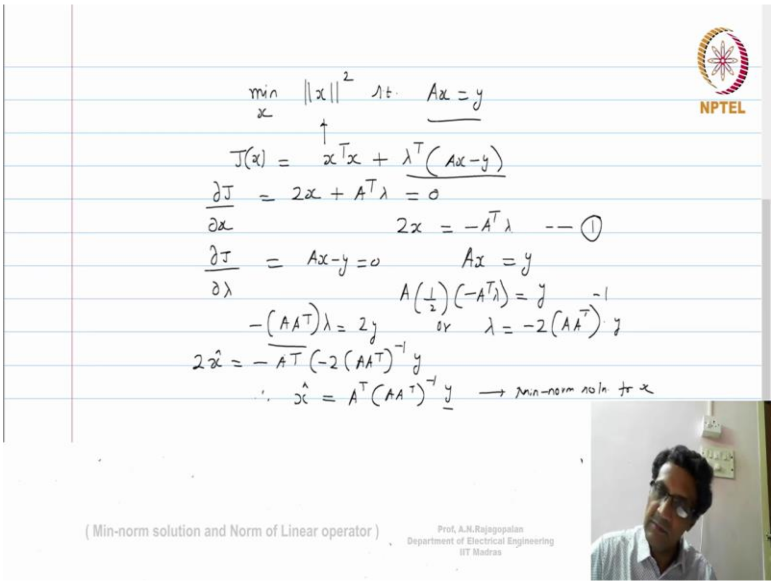
In which case your x hat, then it will then kind of turn out to be h transpose whole inverse h transpose y. That is one way to sort of see how this least square solution will be useful for kind of deblurring problem. But then such a solution is not really the most diligent one to arrive at, we will see how to incorporate for higher and so on. But then I just wanted to indicate that if you have an observation for which a solution does not exist, then what one could do is solve thoroughly least square solution. Now the second position of Hadamard if you recollect, that was about uniqueness. Uniqueness in the sense that if a solution exist, it ought to be unique. This is what it has said.

Now what if, let us say what if you had a situation where let us say y is equal to A x and or let us say A x equal to y which you would even write it as let us say A x equal to y and you want to solve for x where this is again let us say m cross n, this is m cross 1 and this is m cross 1. What if let us say m is less than and, so we have an underdetermined, system of equations and we know that if we have a situation like that, you could end up, you could have multiple solutions for a given y multiple solutions for a given y.

That means that will be multiple x's which will all work for, work to let us say give you this m y. In such a case, one can of course know what can now say, well, in which case I cannot solve this problem because there is no unique x. Although I know that a solution exists, but then there is no unique x that will take me to that y. There are so many x's, which one to pick?

So, what one could do is one could look for this L1 norm for x. So, you could kind of look for that x, which has the smallest L1 norm. This is something sparsity and so on, this is useful in compressor sensing kind of application. But the more easy thing to ask is what is called the min-norm solution. That means find that x from all the x's that satisfy x equal to y, find the one that has the smallest norm. This should be smallest could be that x which has the smallest norm. This is called really the mid-norm solution.

(Refer Slide Time: 3:32)



Now formally to kind of write this, write what we can say is we can look up, so we can say that norm is equal to saying that minimize, so mid-norm solution looks like this. Minimize norm x square because we want this to be minimum subject to the condition that subject to A x equal to y. Because in this case we can satisfy this n equality, not an equality but inequality. Because you know that the solution exist. Therefore, among the x's we want to pick the one that has the smallest norm such that A x that gives you y exactly, not like the least square solution where A x was equal to some y hat, it was not equal to y.

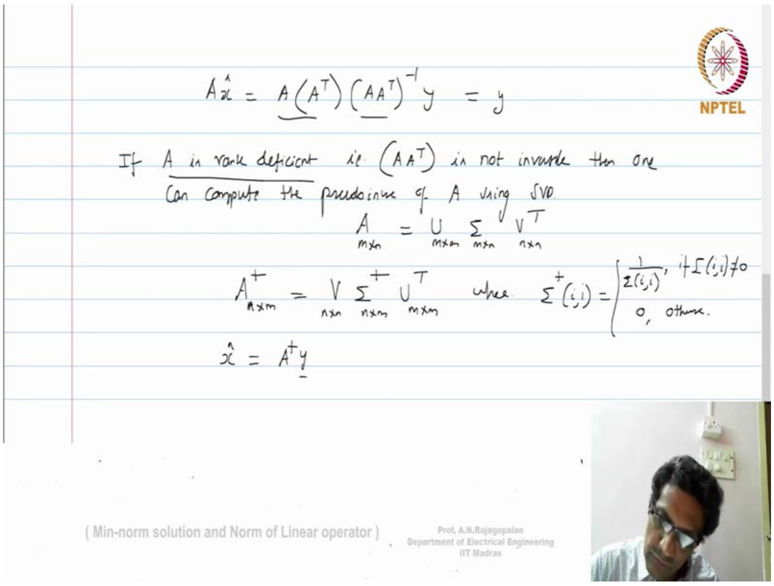
In this case, A x will be equal to y. Now same with in that kind of, now we are now using a Lagrange multiplier, then what you get is you can write this as J of x equal to x transpose x plus a Lagrange multiplier and which you can use here, which is like lambda transpose A x minus y. So, the whole thing boils down to using Lagrange multiplier and it becomes like this. Now you can look at doh J by doh x, that in this case will be 2x plus say y transpose A x, lambda transpose A x is A transpose lambda as I already said.

Even third term of course does not even involve x, therefore this is 0. Or what you can say is 2x equal to minus A transpose lambda, 2x equal to minus A transpose lambda. Now if you try doh J by doh lambda, that will give you simply A x minus y equal 0 because that will be act on this term that will be A x minus y equal to 0. Or in other words, A x equal to y. Now if you substitute for x from this equation 1, then you will get A into x is half minus A transpose lambda, that is x from equation 1, this is equal to y.

Or you can say minus A A transpose lambda is equal to 2y. Or in other words, lambda is you can say minus 2 A A transpose. Lambda is of course not a vector here. Please remember lambda is a vector, so can only, so multiplying A A transpose multiplying the vector from the left, that is why we have A A transpose inverse minus 2 A A transpose inverse y. Now if this is the optimal lambda, put that in here to get your optimal x hat.

Therefore, you will get 2x hat is equal to minus A transpose times, 2x hat is equal to minus A transpose times lambda. So, lambda is minus 2 into A A transpose the whole inverse y. This 2 and 2 will cancel out as minus and minus will cancel out. Therefore, it turns out that x hat is equal to A transpose into A A transpose the whole inverse y. So, carefully if you operate, so this is your mid-norm solution. This is your mid-norm solution for x. Clearly, if you operate A on this x, on this estimate of x, let us operate.

(Refer Slide Time: 6:56)



So, A x hat if you do, you get A into x hat itself. As you found there was A A transpose the whole inverse into A transpose into A A transpose the whole inverse into y. But A A transpose A A transpose inverse is equal to identity. Therefore, this is equal to y, therefore this x hat does satisfy this condition that A x should be equal to 1. So, solution does exist and then among the solution because there are multiple number of them, we are picking the one that has the smallest norm.

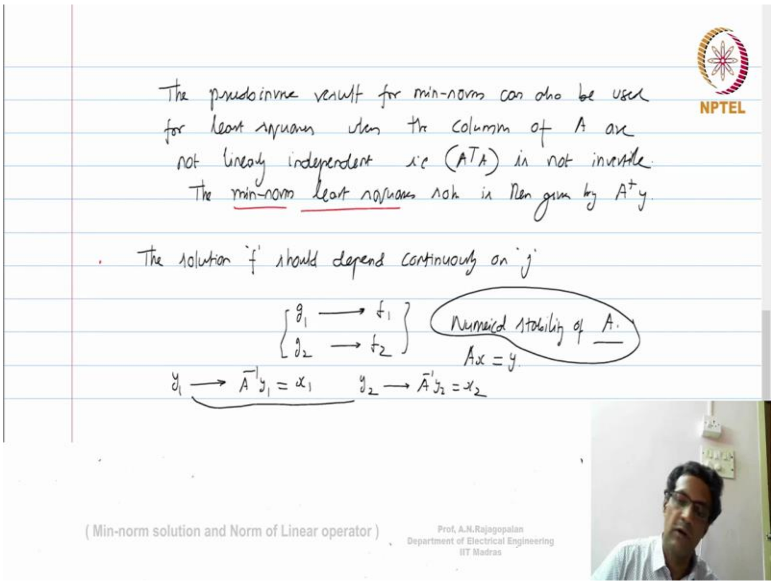
Therefore, your min-norm solution is really this. Now, now if it so happens that let us say x hat is rank deficient, if A is rank deficient, if A is rank deficient, that is A A transpose is not invertible, then one can compute the pseudo inverse and one can compute the pseudo inverse of A using SVD. We have already seen what this SVD is like. So, we can compute the pseudo inverse of A using SVD which then means that given your A which m cross n, you can flick that matrix U which is m cross m sigma U sigma which is m cross m V transpose which is m cross m.

Therefore, A pseudo inverse which is now, it is typically indicated as A plus. So, A plus will be m cross m, this will be equal to V sigma plus, this should be m cross m and then U transpose which would be, so this is n cross m, this is m cross m, where let us write down what we mean by sigma plus and so on. Where sigma plus of I comma i, is equal to 1 by sigma of I comma i. We get this sigma of I comma i, with sigma I comma i is not 0. Else else 0 otherwise.

And that would kind of that give you the pseudo inverse of A and if you operate this A plus and now if you apply this A plus, so the x hat, the min-norm solution will simply be A plus y, A plus acting on y, where A y is the observation. Now this is of course provided that A is, if A was deficient in rank. If not, of course, if not, then you can directly do A A transpose inverse and then while you are getting the min-norm solution.

This is also min-norm by the way. It is also min-norm solution and the same A inverse, let me also mention that now if you go back to your least square solution, we had A transpose A. And if A transpose, you get now, you can even use this result for the min-norm solution, the min-norm result, the pseudo inverse result.

(Refer Slide Time: 10:26)



Result for min-norm can also be used for least squares, can also be used for least squares when the columns of A A are not linearly independent, not linearly independent. What this means is that A transpose A is not invertible in least squares. A transpose A is not invertible in least squares. If this happens, then it means that, that means that the columns of A are not linearly independent.

Then it means that if you try to do least squares, then there is not just one x, there are several x’s that can kind of give you an x hat which when acted upon by A will give you A x hat which is equal to y hat. And that y hat will be as close as possible y in the orthogonal sense. So, that is the error between y and y hat is going to be small. But then among the x’s that will take you there if you use the min-norm idea or if you use the pseudo inverse, then it will give you the x that has the smallest norm.

Even A transpose A is not invertible, of A are not linearly invertible, the columns of A are not linearly, that is A transpose A is not invertible. So, so so the solution, the min-norm, so then such a solution is called min-norm least square solution solution is then given by is then given by A pseudo inverse y. So, it is called the min-norm least square. That mid-norm least squares. So, this is A transpose A is not really invertible.

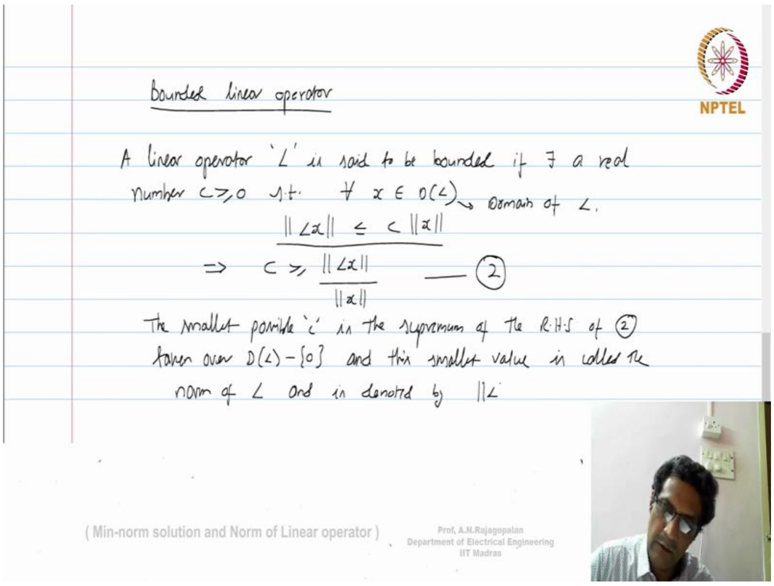
Now these now these two conditions are over as far as we can see. Now what about Hadamard's third condition? Now again min-norm is one way out. If you do not, if the weakness is an issue, least square solution is one way out. Let us say existence is an issue and suppose if somebody told you that let us say I have an operator or I have the metrics that is both one to one I want to, in which case if you have an A which is invertible but you still feel that Hadamard had something else to say, which was his condition number 3.

So, coming to condition number 3 of Hadamard and what he said was he made the right dependence of solution or the solution should depend continuously, solution x should depend continuously on the observation J, should depend continuously on J, should depend continuously on J. Now this is what actually plays the very very important role, where like I said before if G 1, means for G 1 you get a solution f 1 because of the fact that it is invertible, A is invertible.

And G 2 gives you f 2. If G 1 and G 2 are very close then f 1 and f 2, we would ideally want them to be very close. Whether they are close or not depends upon certainly the problem. So, we want to examine the numerical stability of the matrix A, numerical stability, stability fully. So, we are interested in examining what we call is numerical stability of matrix A. So, you are trying to solve A x equal to y and we are saying that y 1, if I have y 1, then my solution is A inverse y 1 which is x 1. And if I give y 2, then my solution is A inverse y 2 which is x 2.

We are saying that if y 1 and y 2 are very close, I expect x 1 and x 2 to be very close. But are they, will they turn out to be close or not depends upon the numerical stability at A. So, we would like to examine this notion in more detail. And in order to understand that we have to talk about norm of an operator and so on which is what we will see next.

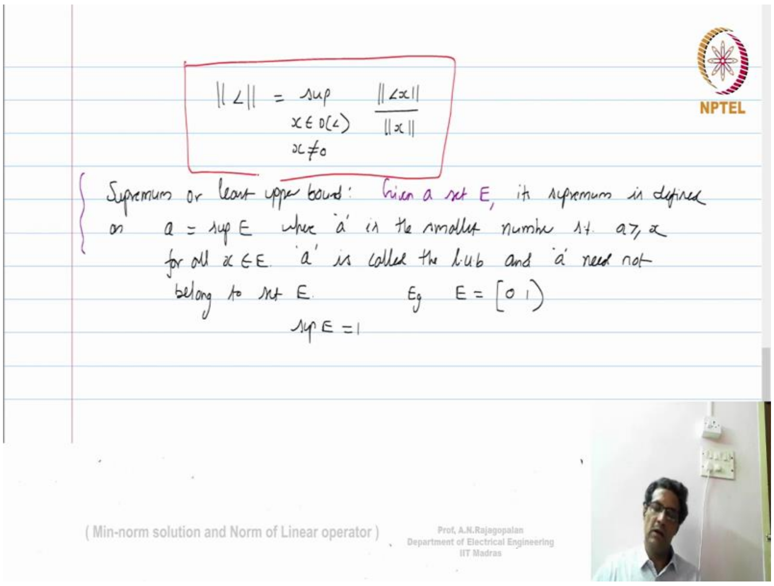
(Refer Slide Time: 15:01)



Let us first talk about a bounded linear operator and then we will kind of see, come down to the matrix. Let us talk about the general notion of a bounded linear operator. A linear operator L is said to be bounded bounded if there exist a real number C greater than or equal to 0 such that for all x belonging to the domain of L, norm L x is less than or equal to C times norm x. This is the domain of L.

This implies that C is greater than or equal to norm L x by norm x. The smallest possible C, so the smallest possible C that satisfies this condition is called the norm of the smallest possible C is the supremum of the RHS of 2, let us say equation 2. Because this is 2 taken over V of L minus the 0 vector. Because since we are dividing by norm x. And this smallest value this smallest value is called the norm of L norm of L and is denoted by and is denoted by norm L norm L.

(Refer Slide Time: 17:25)



That is norm of L is equal to supermom over all x belong to the domain L, x naught equal to 0 of the quantity which is given as norm L x by norm x. Now when we write supermom, we should understand the supermom is not the same as maximum that is, slightly more kind of general idea. So, what do we mean by for those of you, one thing what supermom is or what is called the least upper bound?

Is called the least upper bound. So, this is simply aside for those of you, so given, for example, given a set E, within a set E its supermom is defined as a is equal to sup of E where a is the smallest smallest number such that a is greater than or equal to x for all x in E, for all x belonging to E. Important to note that the, the important thing to note is that a is called the Lub or supermom or whatever, least upper bound and a need not belong to set E.

Whereas, whenever you talk about maximum, it comes from the set. When it is a maximum of the set, then you actually pick a number that actually belongs to that set. Whereas whereas when you talk about supermom, it does not even have to belong to that set. For example, if you take E to be the slow set from the left and open it from the right, then the super V is equal to 1 and you see that 1 is not in the set E.

But then we know that 1 is the smallest number that is greater than or equal to all numbers in the set E. But then set E does not include 1 by itself because because this is an open interval. That is the notion of a supermom. Now using this notion we would like to understand how to now talk about the norm of the matrix. Because this is a definition for norm of linear operator and therefore, one would like to use this kind of definition in order to be able to talk about norm of a matrix.