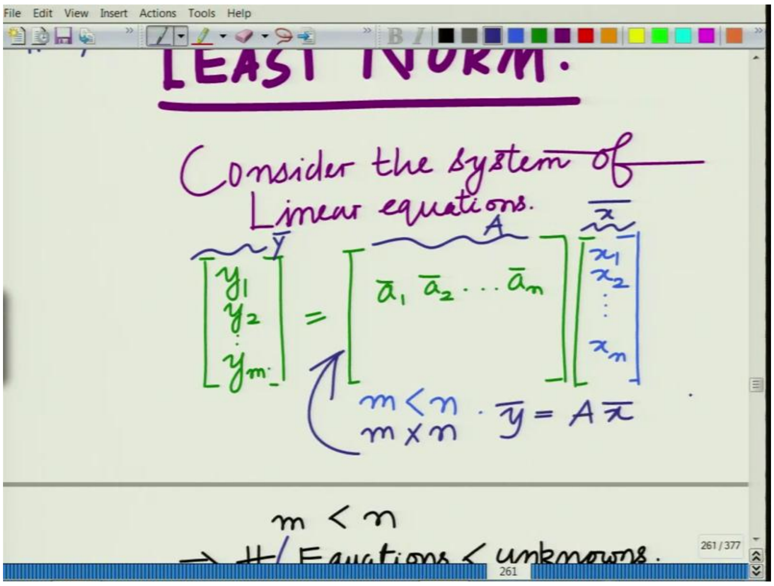
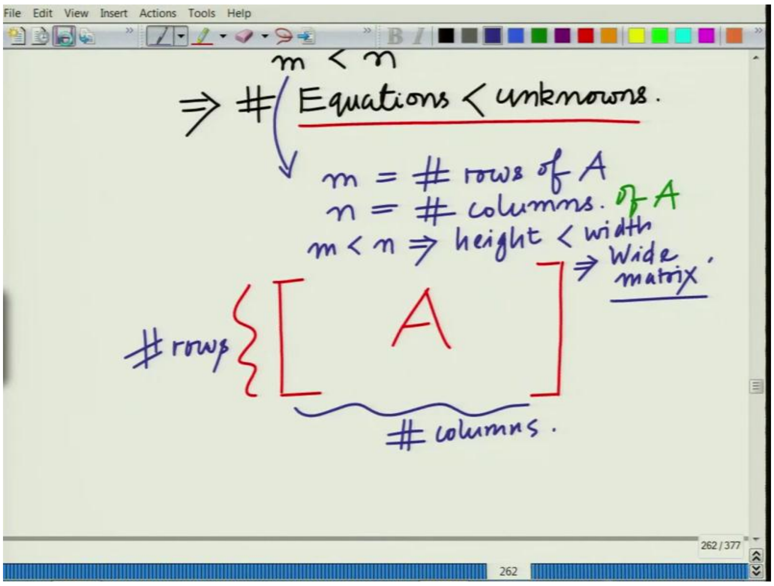
Applied Linear Algebra for Vector Processing, Data Analytics and Machine Learning
Professor Aditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Lecture 27
Least Norm Solution

Hello. Welcome to another module in this massive open online course. In this module let us start looking at another important concept in linear algebra, and also pertaining to the solution of system of linear equations and that is the least norm solution. So, far we have seen the least square solution, we will now start exploring the least norm solution. So, what we want to do is essentially we want to look at the least norm solution.

(Refer Slide Time: 1:16)





So, this is the least norm. Previously we have looked at the least squares. Now, we want to look at the least norm solution. So, once again, consider the system of linear equations. Consider the system of linear equations, we have y1, y2, ym this is equal to once again the matrix A, comprising of the columns a1 bar, a2 bar so on up to an bar times the vector x bar, which is x1, x2, up to xn plus that is it, this a system of linear equations.

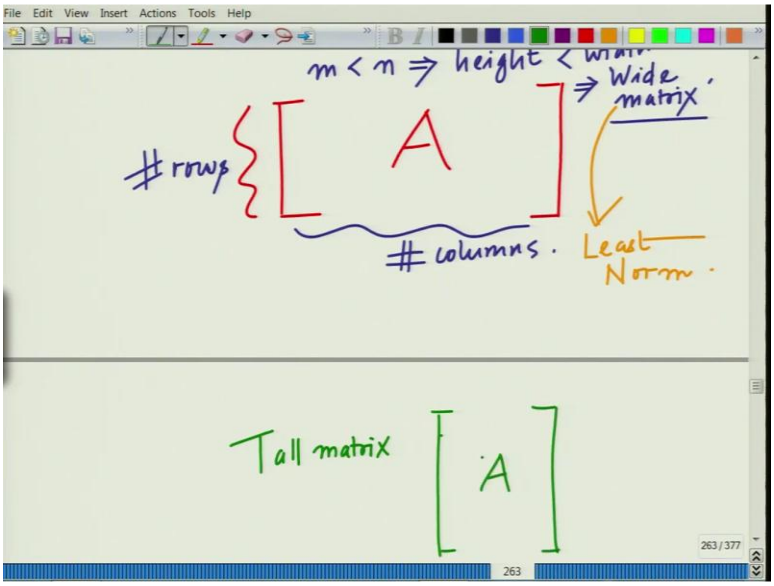
Except in this case, so we have this matrix this is an m cross n and for the least norm problem specifically, we have m less than n. So, we have m cross n matrix, m is the number of equations, n is the number of unknowns and we have m less than n. So, this system of linear equations is essentially characterized by the fact that number of equations. So, m less than n, this implies that number of equations is less than number of unknowns.

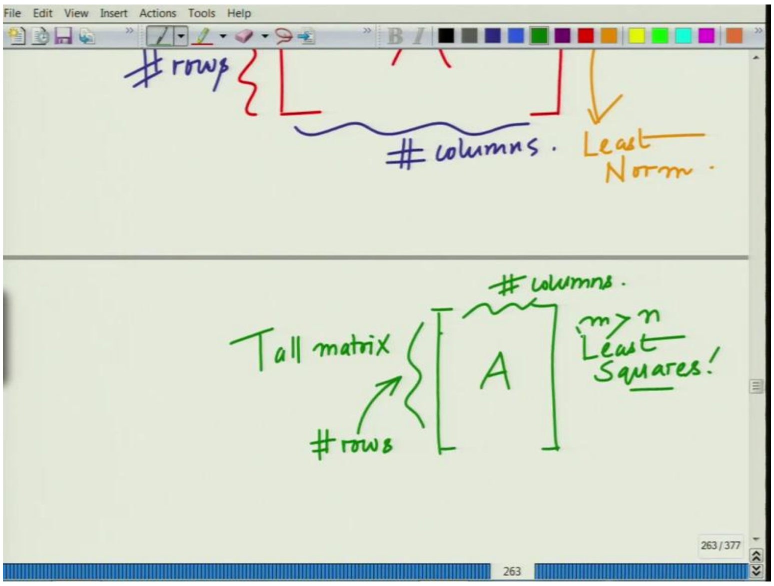
So, the number of equations is less than the number of unknowns. Also m less than n remember, m equal to number of rows of the matrix A. So, you have this is basically your matrix y bar, this is basically your matrix A, and this is your matrix x bar, so you have y bar equals A times x bar. So, you have the matrix A in equals number of columns of A, of the matrix A. So, if we look at the matrix A, that looks like, so this is your matrix A, this is the number of rows and this is the number of columns.

So, number of rows is less than the number of columns, this implies m less than n implies that the height of the matrix. If you look at the number of rows as the height, so the height of the matrix is less than the width implies this is a wide. Implies, this is what is known as a wide matrix. So, if you look at, if you want to find a, form a picture of the system in your mind, it is always useful to form a mental picture in your mind. So, we have an underdetermined system of linear equations.

So, to form a mental picture, essentially I have a matrix A, in which the number of rows is less than the number of columns that is the width, that is the height is less than the width, which implies this looks like a wide matrix essentially. Remember the least squares was the other way around, the least squares had number of equations more than the number of unknowns. So, number of rows was more than the number of columns overdetermined system of matrix, overdetermined system of linear equations and therefore, the corresponding matrix looks like a tall matrix.

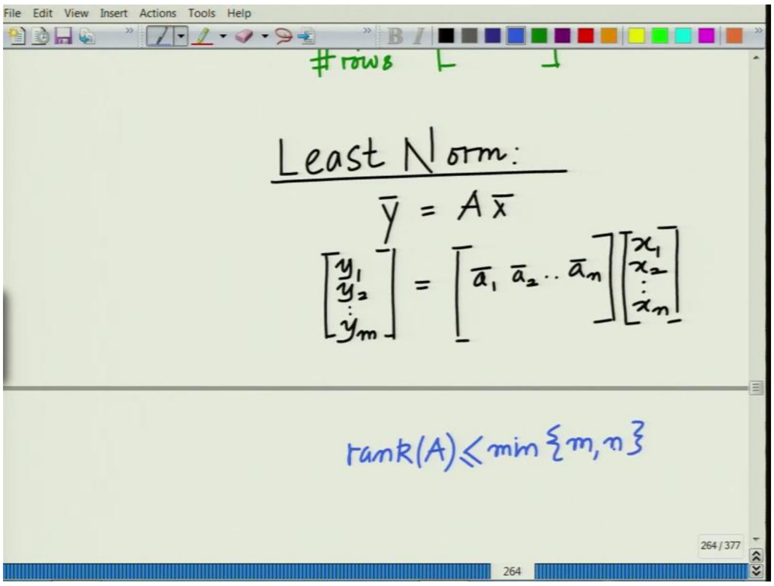
(Refer Slide Time: 6:19)





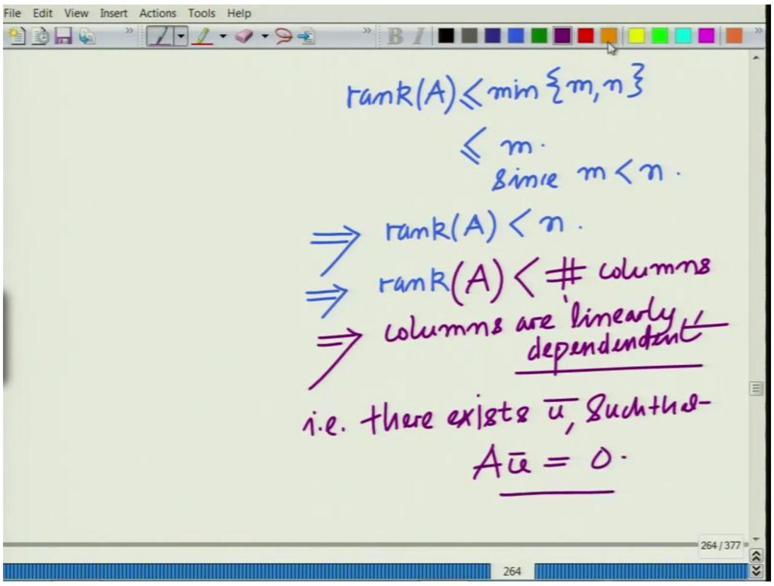
So, the least norm, so the wide matrix is for the least norm and the tall matrix if you look at it, in which you have the number of rows and this you have the number of columns in which m less than n, this is for the least, this is for your least squares. So, just remember, it is useful to remember that least squares is usually associate with a tall matrix, least norm is associated with a wide matrix.

(Refer Slide Time: 7:29)



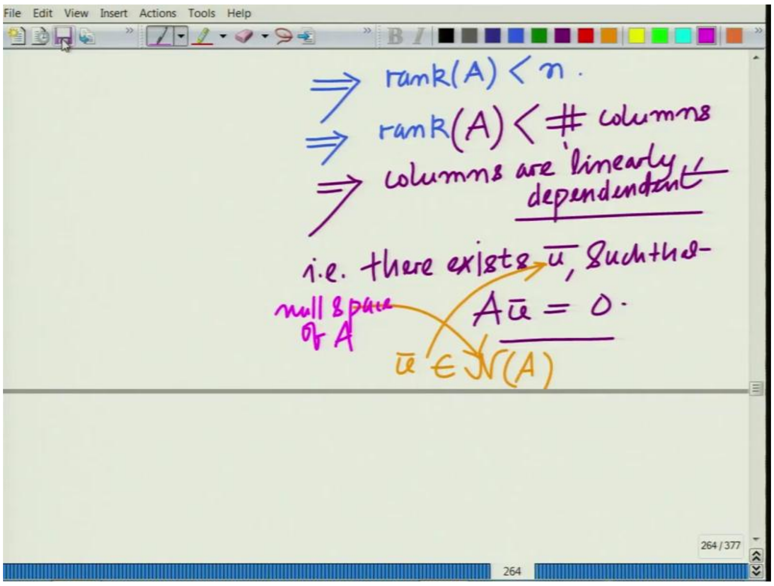
Now, so coming back to our least norm, what you can see is, now you have y1, y2, ym equals your matrix A, A1 bar, A2 bar, an bar, x1, x2. Now, basically if you look at the rank of this matrix, rank A, if you look at the rank of this matrix, which is less than or equal to the minimum of m, n, number of rows comma number of columns.

(Refer Slide Time: 8:52)



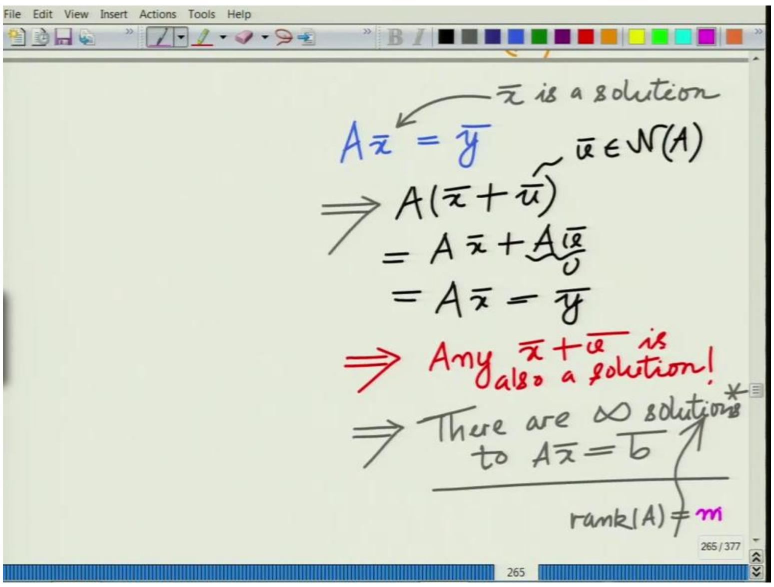
Which now m, given m less than n, this implies this is less than or equal to m, since we have m less than n. So, rank is less than or equal to m, which implies the rank of A is less than n, which essentially implies that rank of A is less than the number of columns. So, implies the columns are linearly dependent or not implies the columns are linearly dependent. That is, there exists a vector u bar such that a times u bar equal to, so that is, u bar, such that A times u bar is 0.

(Refer Slide Time: 10:21)



And in fact, we said, we defined such a vector u bar, this u bar belongs to the null space, this u bar belongs to the null space of A. So, this is basically your null space of u bar, belongs to the null space of the matrix A, which means that this wide matrix essentially has a non-trivial null space. So, for instance if you have a single vector, u bar, which belongs to the null space, then any multiple of u bar that is K times u bar also belongs to the null space and so on and so forth and then you will have a basis for the null space and so on and so forth.

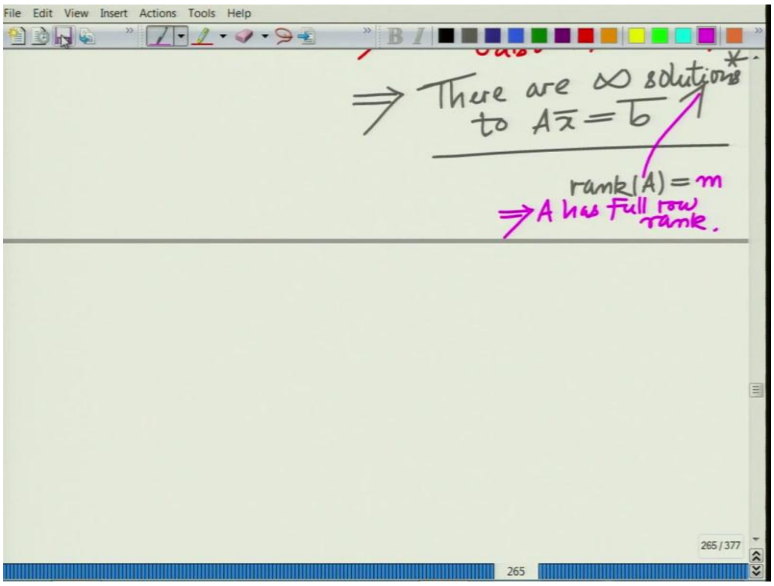
(Refer Slide Time: 11:08)



So, the bottom line here is that, if you look at any solution, let us say you have Ax bar equal to y bar which is a solution. So, let us say x bar satisfies this. So, x bar is a solution, if x bar, if x bar satisfies this implies any Ax bar plus u bar is equal to Ax bar plus Au bar. So u bar belongs to the null space of A. Now, Au bar is 0, so this is equal to Ax bar which is equal to y bar. And therefore, this implies that, if x bar is a solution then any Ax bar, then any x bar plus u bar is also a solution then any x bar plus u bar is also a solution, implies there are, there are infinite number of solutions, infinite, there are infinite solutions to Ax bar equal to b bar.

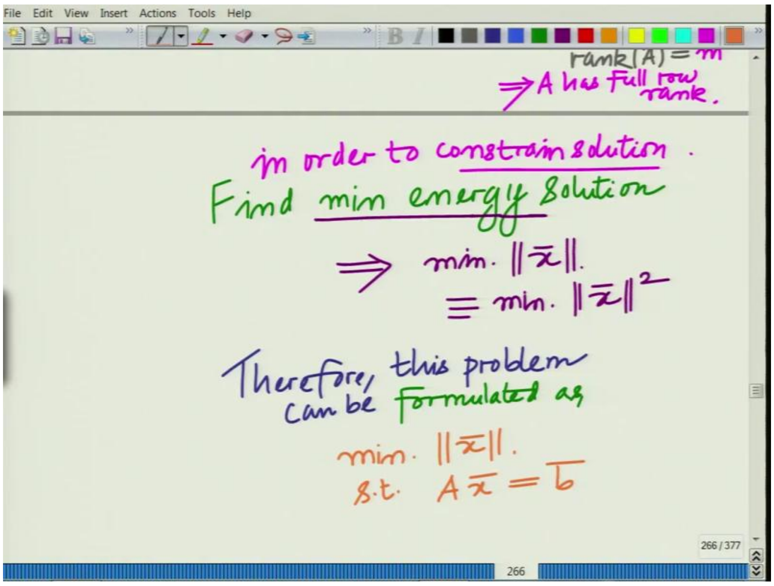
The number of solutions to this system is infinity. So, in fact, if you compare it once again with the least squares, so when you have a tall matrix, the problem there is that there is typically no solution. The solution exists only if y bar lies in the column space of the matrix A. So, in the over determined case, there is no solution. In underdetermined case, typically, there is an infinite number of solution. Of course, once again we have to clarify, there is a slight caveat, if A has full row rank, that is if the rank, that is for any vector A this will have an infinite number of solution, I will just qualify this there are infinite solutions. The qualification here is if once again the rank of A is equal to m.

(Refer Slide Time: 14:14)



That is rank of A equal to m meaning that is implies A has A has full row rank. If rank of A is m, implies A has full, in there have been, in that situation this has infinite number of solutions, which means now we have to constrain the solution. So, we want to find a particular solution. Now, typically given the infinite number of solutions, how do we isolate the best solution? One of the ways to isolate the best solution is typically, especially in machine learning signal processing, it is usually seen that the best solution among these infinite number of the solution is one that has the minimum energy.

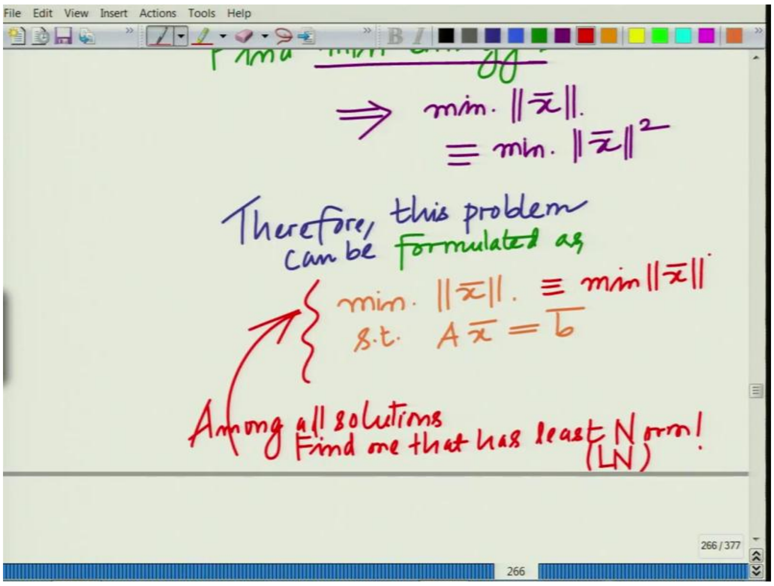
(Refer Slide Time: 15:07)



So, we want to find the minimum energy solution. Find the minimum energy solution. Implies minimize, remember, we said the minimum, if you represent a signal by a vector then the minimum is the energy is given by the norm. So, we minimize norm x bar, which again is similar to minimizing norm of x bar square. Because remember, if we are minimizing norm of x bar that is equal to minimizing the norm of x bar square. And this we are doing in order to, in order to constrain the solution. What do we mean by that?

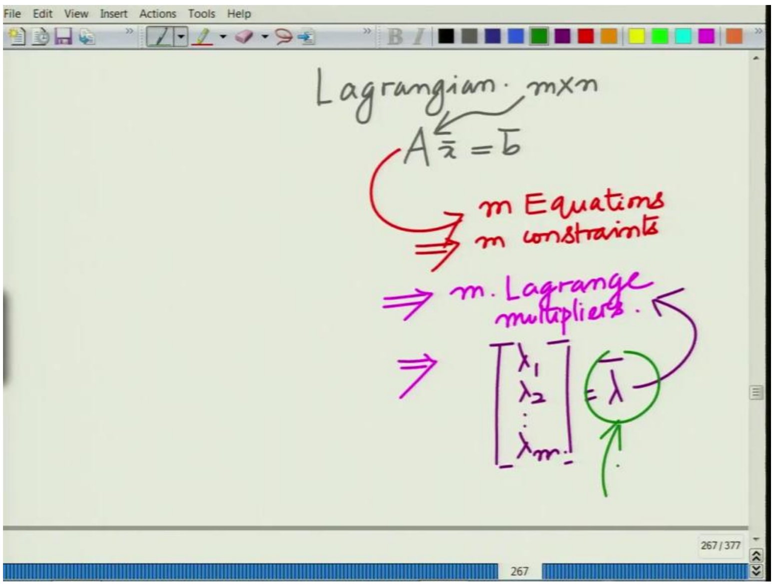
Because there is an infinite number of solutions, so you cannot determine a unique solution, therefore we have to impose additional constraints. One of the logical constraints that we are saying and practical useful constraint is to determine the signal or the classifier, which has the minimum energy and essentially that is what we are trying to do here, minimizing the norm, find the solution x bar, which has the minimum norm. And therefore, this problem can be formulated as minimize the norm of x bar subject to the constraint Ax bar equal to b bar.

(Refer Slide Time: 17:32)



So, what, what we mean by this is among all the solutions, find the one that has the minimum energy or the minimum norm or the least norm. Therefore, this is known as the least norm solution. Remember, previously, we had the least square solution, now we have the least norm solution. So, among all the solutions, find the one that has the least norm. This is the least norm solution. Among the infinite number of solutions, find the one that has the least norm, that is why this is the least norm problem. And remember, you can also equivalently formulate this as minimize norm of x bar square.

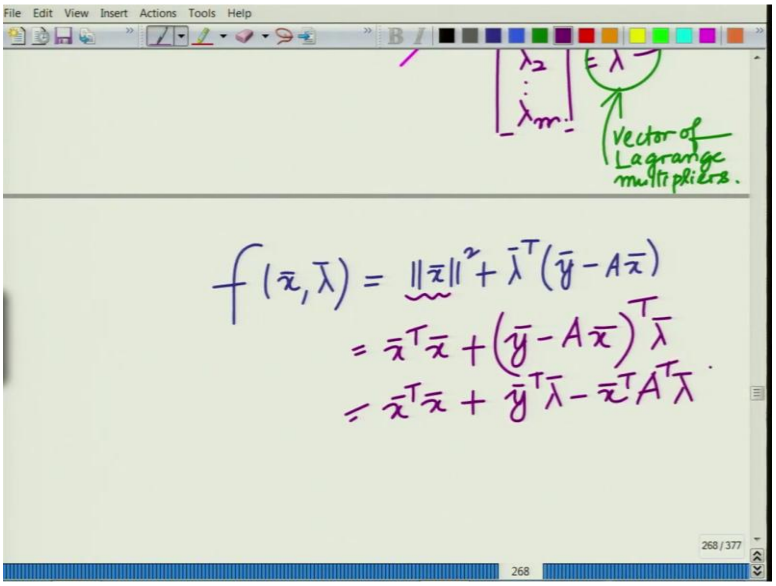
(Refer Slide Time: 18:48)



Once again this can be solved using the Lagrangian Convex. So, this is a convex problem this can be solved using the Lagrangian cost function. Except here you realize that you have Ax bar equal to b bar, which means A is m cross n there are m equations implies there are m constraints.

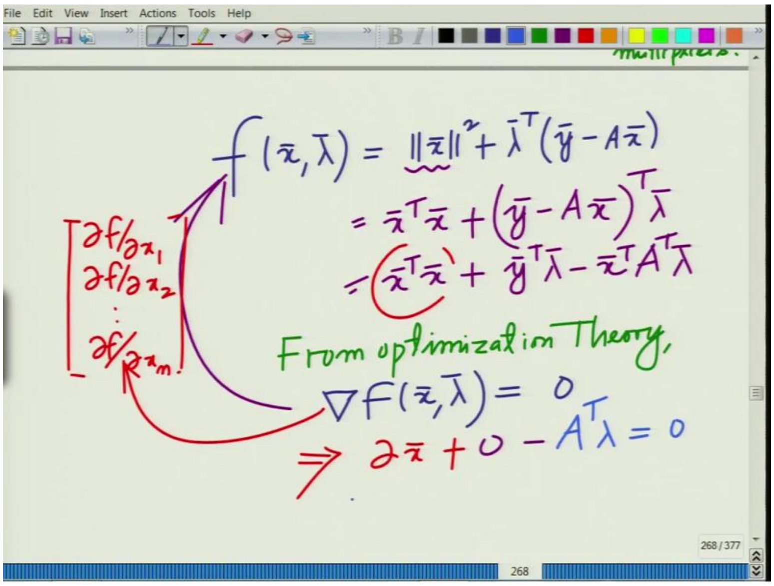
And therefore, we have, we need m Lagrangian, some of you might be familiar with this theory, which implies that you need m, which implies that you will have the vector lambda 1, lambda 2, lambda n this is equal to lambda bar, this is the vector Lagrange multipliers. So, we can say this is the vector of Lagrange multipliers, because you have m constraints.

(Refer Slide Time: 20:12)



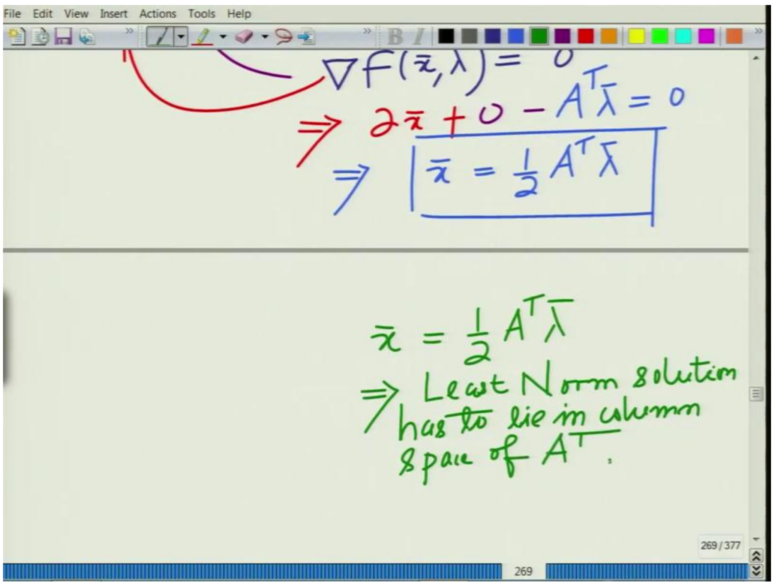
So, this is the vector of Lagrange multipliers. And therefore, you can formulate the Lagrangian for this as f of x bar comma lambda bar that will be minimize norm of x bar square plus the Lagrange multiplier vector lambda bar transpose times y bar minus Ax bar, norm of x bar square, I can write this as x bar transpose x bar plus I can write this as y bar minus Ax bar transpose lambda bar. A transpose b bar equal to b transpose A bar, which I can write as x bar transpose x bar plus y bar transpose lambda bar minus x bar transpose A transpose lambda bar.

(Refer Slide Time: 21:37)



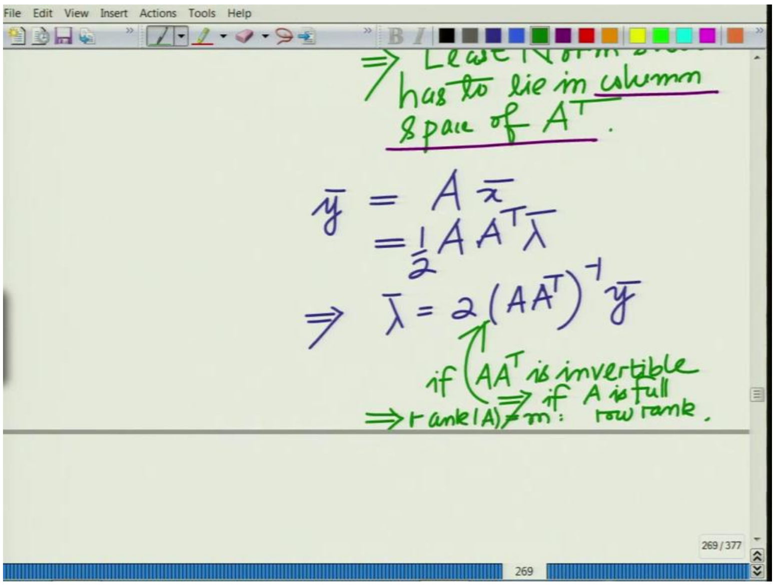
Now, again from the optimization theory, at optimum we must have the gradient with respect to x bar this must vanish. Now taking the gradient, we have already seen that gradient of, what is the gradient again you might recall the gradient is nothing but the vector of partial derivatives take the derivative of f with respect to each component of x. So, this implies that x bar transpose x bar, the derivative is twice x bar plus y bar transpose lambda bar derivative with respect x bar is 0 minus x bar transpose A lambda, x bar transpose A transpose lambda bar derivative with respect to x bar is A transpose lambda bar equal to 0.

(Refer Slide Time: 22:54)



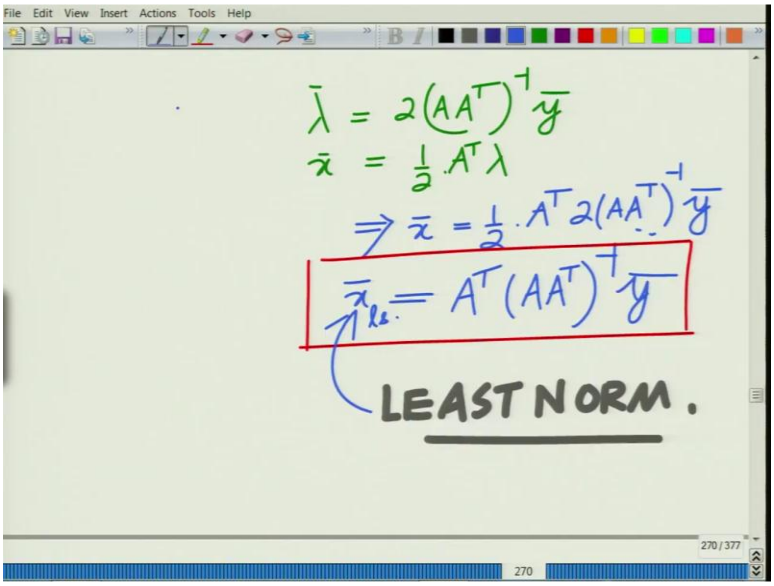
Which essentially implies that x bar is equal to half A transpose lambda bar. This is the property of the optimal solution least norm solution. And you can see a very interesting property if x bar, optimal x bar lies in the column space of A transpose, this is a very interesting property, the optimal x bar, x bar equal to A transpose lambda bar this shows that the optimal x bar has to lie in the column space of A transpose. So, x bar equal to half A transpose lambda bar implies the least norm solution has to lie in column space. of A transpose.

(Refer Slide Time: 24:06)



This is the interesting solution, column space of A transpose. And now, substitute this property in the constraint, we have y bar equals Ax bar equals A times half A transpose lambda bar is implies lambda bar equals twice AA transpose inverse y bar, which is possible if AA transpose is invertible, which once again is possible if A is full row rank. Remember for the least squares A transpose A is invertible, if A is full column rank. Here AA transpose is invertible if A is full row rank, that is rank of A equal to m. It implies rank A equal m.

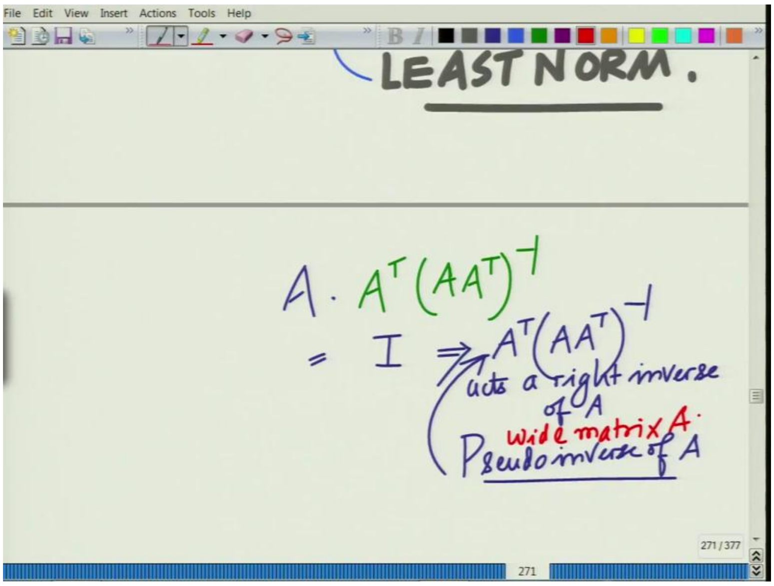
(Refer Slide Time: 25:28)

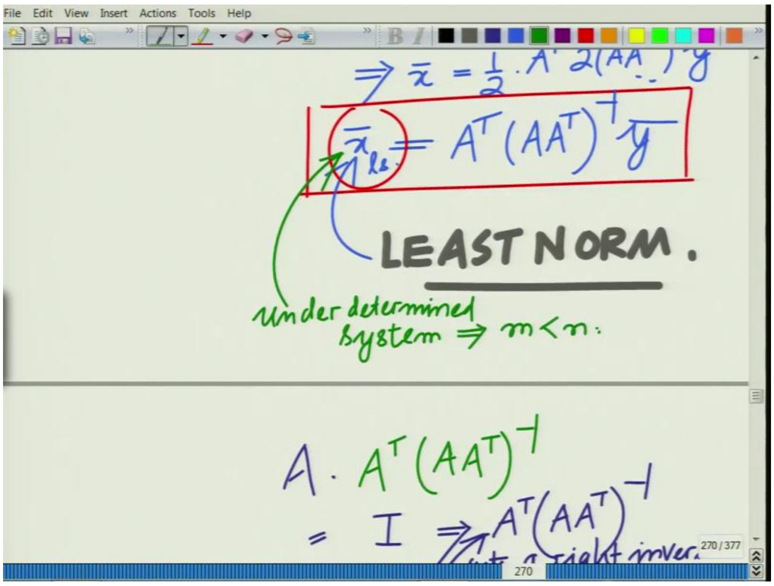


And now, substituting this, so we have these two properties, so we have lambda bar equals twice AA transpose inverse y bar, and we have x bar equal to half A transpose A transpose lambda bar. Now substituting the expression for lambda bar this implies x bar equal to half A transpose, substitute for lambda bar twice AA transpose inverse y bar which is essentially equal to A transpose AA transpose inverse y bar, this is your least norm. And we can call this as x bar of LS.

This is basically what we call as the least, let me just emphasize this, this is your least one of the other the counterpart of the least squares for underdetermined and this is the least norm, both these again have several applications and these are very useful to remember, the least squares and least norm. And you can see they are very related, least squares for overdetermined, least norm for under determined. And this is given by A transpose AA transpose inverse y bar, if A has full row rank.

(Refer Slide Time: 27:10)





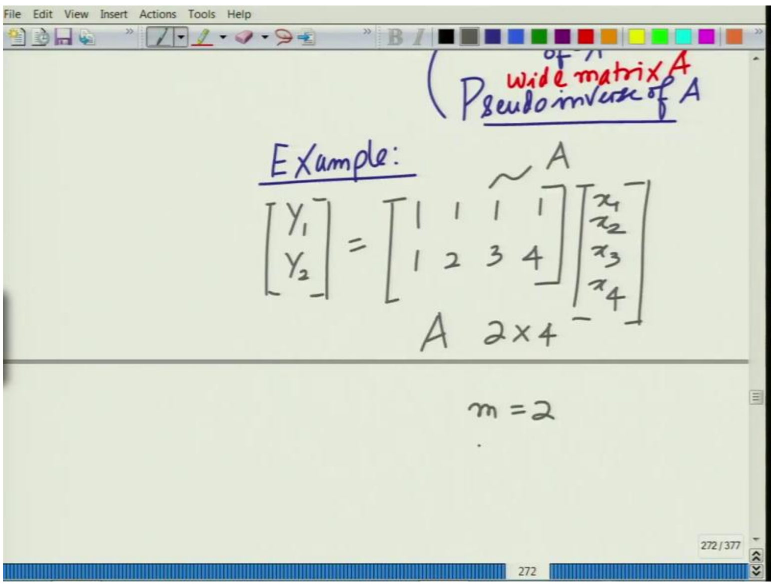
Once again you can see the interesting property, which is if you look at this matrix, A transpose, AA transpose inverse, then if you multiply this on the right with A, this is equal to A transpose into AA transpose inverse this is equal to identity, which means A transpose AA transpose inverse acts as right inverse of A and this is termed as the pseudo inverse.

We have seen the concept of pseudo inverse, but earlier we had seen the concept of pseudo inverse for a tall matrix that is given by A times A transpose A inverse and you multiply it on the left. Now, for a wide matrix the pseudo inverse is A transpose AA transpose inverse and this is basically a right inverse. So, this is a right inverse. So, in the wide case of a wide matrix, this is the pseudo inverse is given by A transpose AA transpose inverse. And this is basically you, are right inverse.

So, this is pseudo inverse of A for a wide matrix. It is important to call it. Because you might see two definitions of the pseudo inverse and might get confused. There is no need to get confused, the earlier definition was for a tall matrix, this is for wide matrix. And the other interesting thing is A square, if A square is invertible, left inverse right inverse both reduce to the same that is A inverse and therefore you will have A inverse A equal to AA inverse equal to identity. So, for a square matrix both of these reduced to the same which is again A inverse, and that is also not very difficult to see.

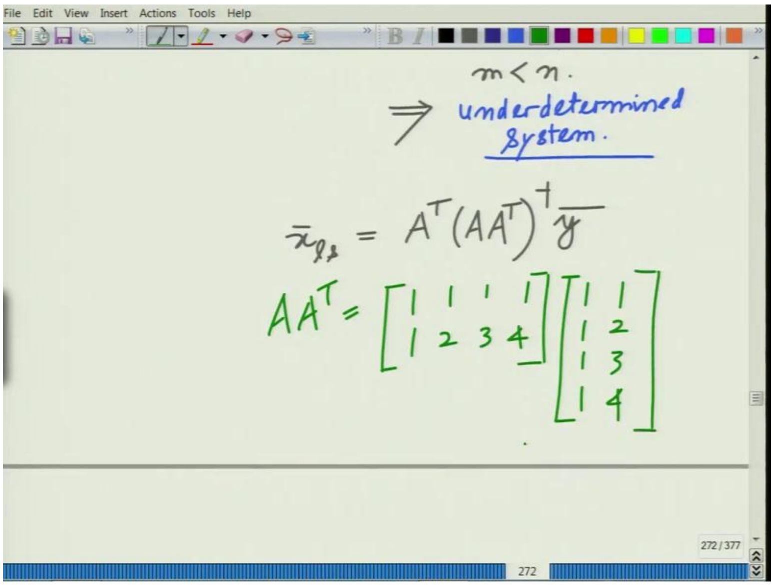
So, this is the least norm solution and this is essentially the right inverse. So, this is again another very interesting and it has several applications, so you have the least norm solution it is, the analog as I already told you the analog of the least square solution for a underdetermined system of equations. So, let us use, also good to remember that this is for a underdetermined, underdetermined system that is m less than n. Because again, once again remember that this is because you have an infinite number of solutions and therefore you want to constrain the system. Let us look at a quick example.

(Refer Slide Time: 30:10)



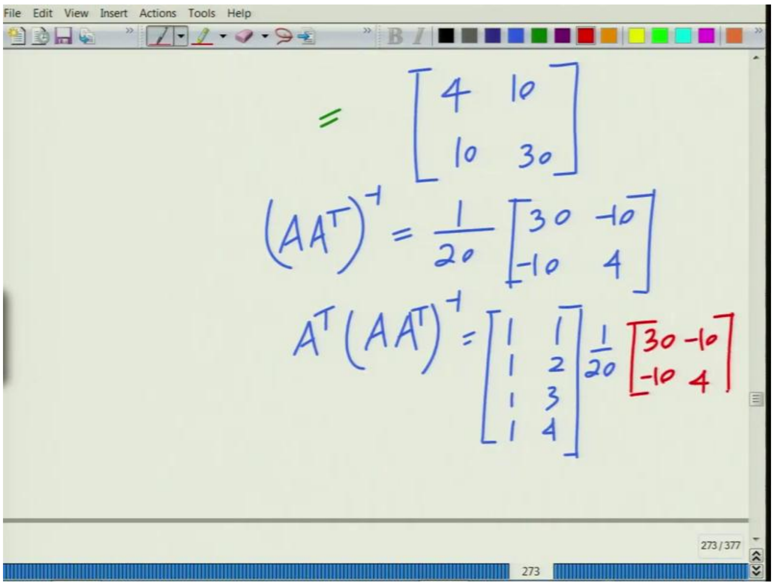
Once again, quick example. Always useful to do examples, so let us have a quick example, y1, y2 equals 1, 1, 1, 1 our favourite matrix 1, 2, 3, 4 and then you have x1, x2, x3, x4. So, this is your matrix A, you can clearly see A is 2 cross 4, m equal to 2, n equal to 4, m less than n.

(Refer Slide Time: 31:00)



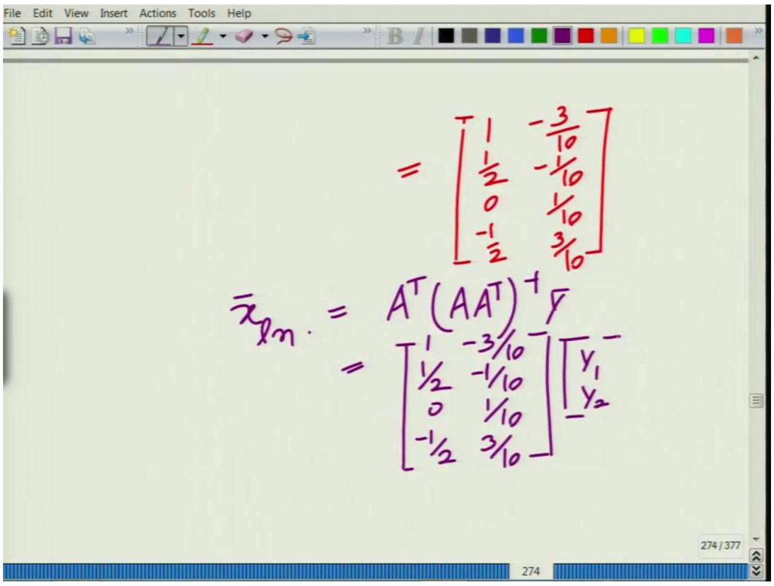
Which implies this is an underdetermined solution. Therefore, one can again find the least norm solution. The least norm solution is given as A transpose AA transpose inverse y bar. Let us ask the first question what is AA transpose inverse. Let us find first AA transpose. So, that will be 1, 1, 1, 1; 1, 2, 3, 4 times 1, 1, 1, 1; 1, 2, 3, 4.

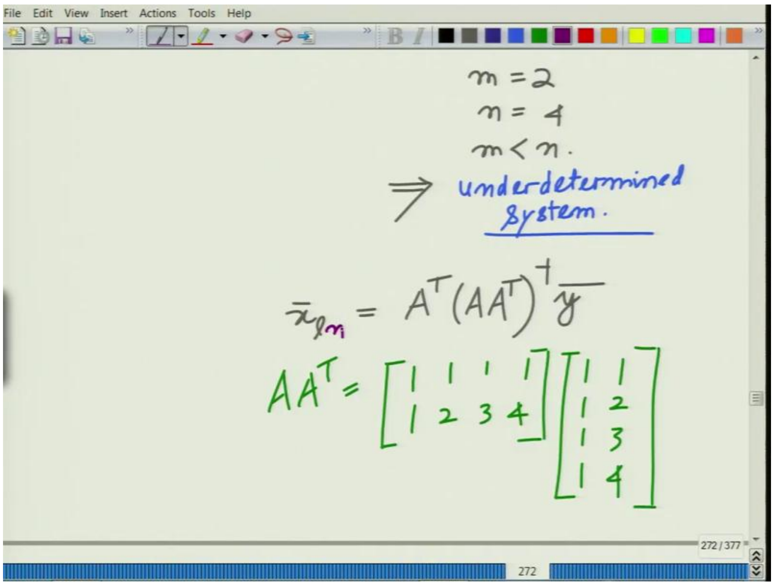
(Refer Slide Time: 31:56)

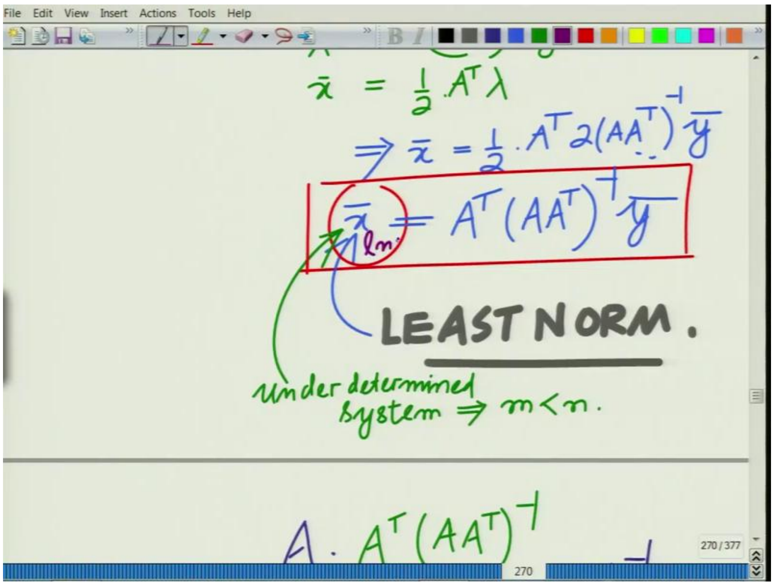


Which if you look at this as given by this matrix this will be 4, this will be 30, this will be 4. The cross elements will be 4, plus 3, 7 plus 2, 9 plus 1, 10 and AA transpose inverse equals 1 over 120 minus 100 equals 20 times, 30 4, minus 10, minus 10. And once again if you calculate the pseudo inverse that is A transpose AA transpose inverse which is 1, 1, 1, 1; 1, 2, 3, 4 times 1 over 20 into the matrix 30, minus 10, minus 10, 4.

(Refer Slide Time: 32:59)



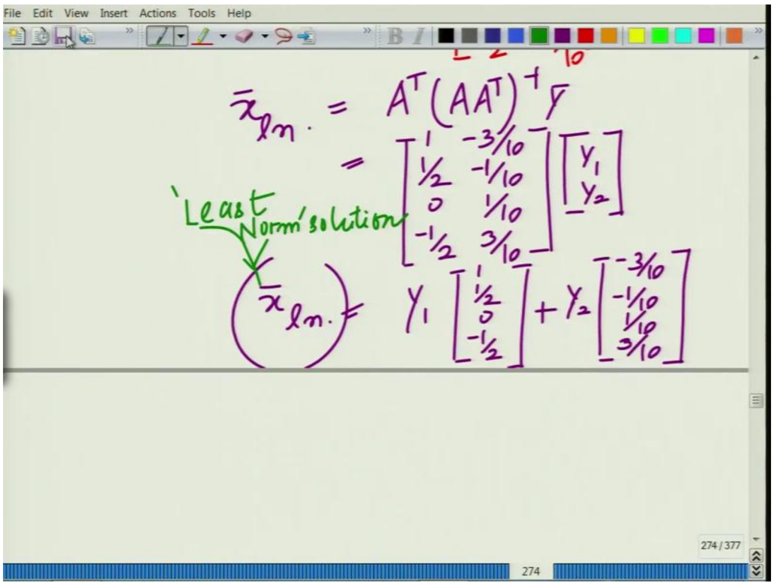




You can evaluate this again this will be equal to 1, half, 0, minus half, minus 3 by 10, 1 by 10, 1 by 10, 3 by 10. That is 3 by 10, minus 1 by 10, 1 by 10. So, this will be 1, half, 0, minus half; minus 3 by 10, minus 1 by 10, 1 by 10, 3 by 10. And therefore, the least norm the least norm solution is given as x bar least norm, sorry, I think I have been writing ls everywhere, I think this has to be x bar least norm.

So, this is your x bar least norm, x bar ln, x bar ls you can think of it as the least squares solution. This is your x bar ln, that is the least norm solution. So, I think that makes sense. So, x bar ln equals the least norm solution equals this matrix that is your A transpose AA transpose inverse y bar which is 1, half, 0, minus half; minus 3 by 10, minus 1 by 10, 1 by 10, 3 by 10 times y1, y2.

(Refer Slide Time: 35:04)



So, this will be equal to well, y1 times the first column of this matrix 1, half, 0, minus half plus y2 times minus 3 by 10, minus 1 by 10, 1 by 10, 3 by 10, and this is your least norm solution to that given system. So, this is basically your least norm solution. So, this is the principle of least norm solution, this is the theory of least norm solution. And as I told you this is again another very important concept of linear algebra, least squares, least norm, sort of two sides of the same coin. And this together constitute now the complete set of possible solution for a linear system of equations.

So, linear a system of equations m equal to n number of equations is the number of unknowns, if A is invertible then you have a unique solution x bar equal to A inverse y bar. If m is greater than n, number of equation is greater than number of unknowns, then you have an overdetermined system of equations and which you use the least squares, and if you have m less than n, which is now the case that we have seen now, you have an underdetermined system of equations in which we use the least norm.

So, together, this constitute the complete spectrum of solutions for the linear system of equations. Of course, there are minor cases where you can have things such as for instance, A not being full column rank, A not being full row rank and then in that case, you will have to the solution becomes a little more tricky, but those are special cases what I would like to call us pathological cases. But by and large this the unique solution given by the inverse least squares, least norm these constitute the complete spectrum of solutions to a linear system of equations following by enlarge.

And these are important to know because these arise very, very frequently in practice. And these are again the one of the most significant applications of the principles of concepts of linear algebra that we have learned so far. So, let us stop here and we will continue in the subsequent videos. Thank you very much.