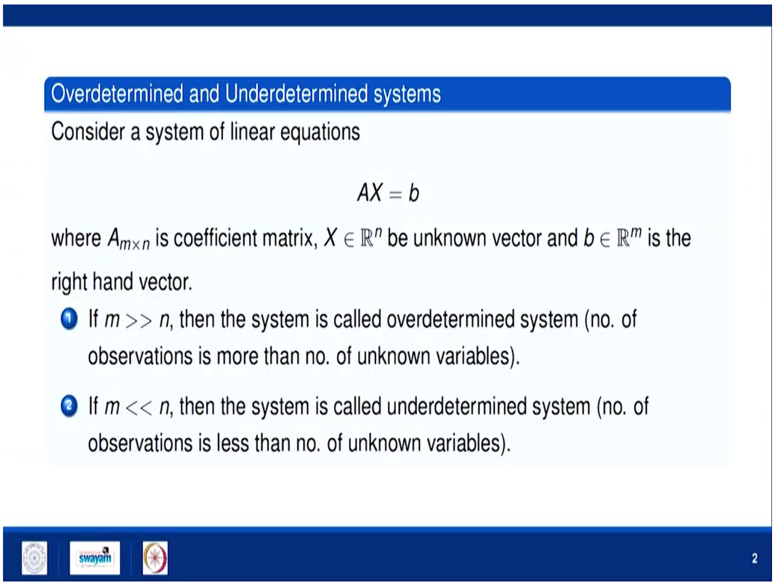
Essential Mathematics for Machine Learning

Prof. Sanjeev Kumar
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture – 10
Least Square Approximation and Minimum Normed Solution

Hello friends. So, welcome to the module 10 of this course Essential Mathematics for Machine Learning. So, in this lecture we will talk about a very important concept of solving linear system of equations, which frequently occurs when we train our system in supervised learning. So, we will talk about Least Square Approximation and Minimum Normed Solution.

(Refer Slide Time: 00:50)



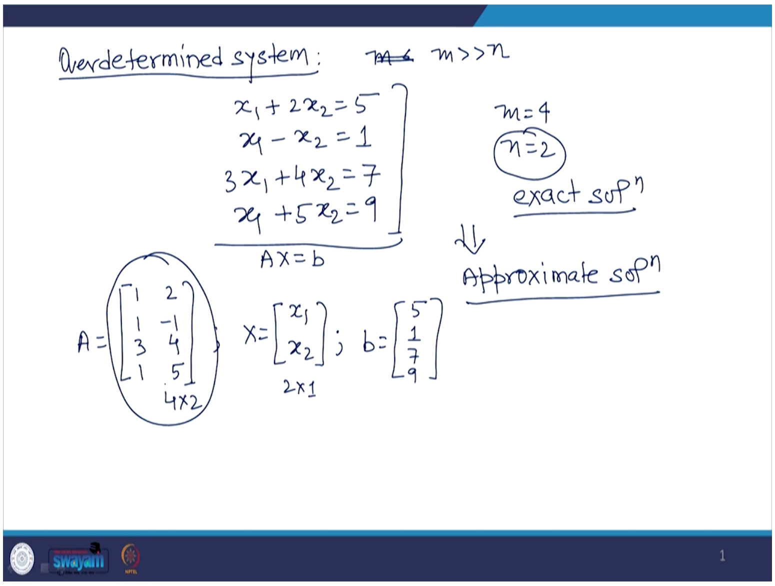
So, consider a linear system of equations A X equals to b, where A is m by n coefficient matrix, X is the unknown vector belongs to R n and b belongs to R m is the right hand side vector, which is given to us.

So, here just notice we are taking a rectangular system A, A is a m by n matrix. If it is a square matrix let us say m equals to n. So, m by n matrix and it is invertible then the solution will become X equals to A inverse b. But here we are talking when A is not a square matrix.

So, if m is bigger than n, let us say m equals to 100 and n equals to anything less than 100, then the system is called over determined system. Because here number of rows are more than number of columns. So, what is the meaning of number of rows, that we are having more observations when compared to the number of unknowns.

In this case we say that system is a over determined system. Similarly when m is less than n means we are having more number of unknown variables when compared to the number of equations, then the system is called underdetermined system. So, what happen in case of over determined system?

(Refer Slide Time: 02:24)



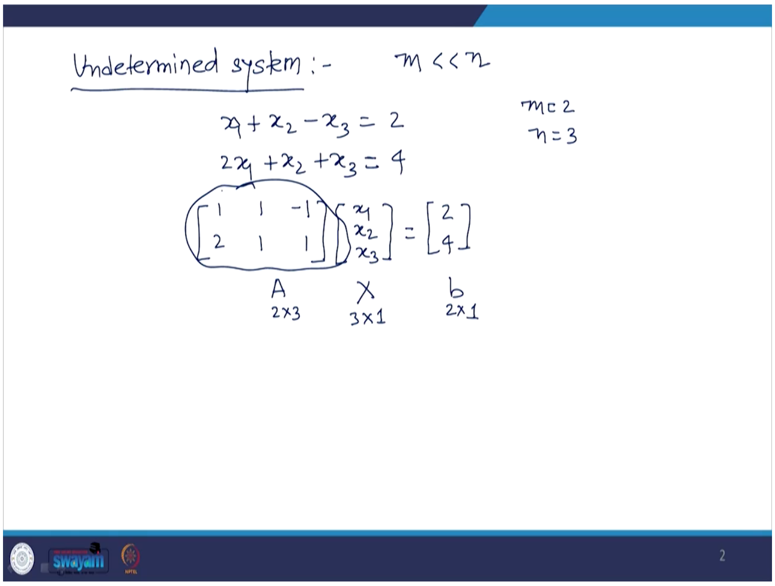
So, in over determined system we are having m is bigger than n, means more number of rows. So, for example, you consider an example of like this, something like this. So, here we are having m equals to 4 and n equals to 2.

So, four equations are given to us, based on these four equations we have to determine x 1 and x 2. So, in this case what will happen? We are having an over determined system and exact solution will come very rarely.

So, what we have to look? We have to look for n approximate solution. Such an approximate solution is called least square approximation of over determined system. So, how can I write this in matrix form? So, I can write it as A X equals to b, where A is 1 2, 1 minus 1 coming from here, 3 4 coming from here, 1 and 5. Similarly X will become a known vector.

So, x is x 1 and x 2. So, it is 4 cross 2, it is 2 cross 1 and what it might be, b is given right hand side vector. So, 5, 1, 7 and 9 so here in case of over determined system, the matrix A which is the coefficient matrix is a silly matrix, that we are having more number of rows then column.

(Refer Slide Time: 04:59)



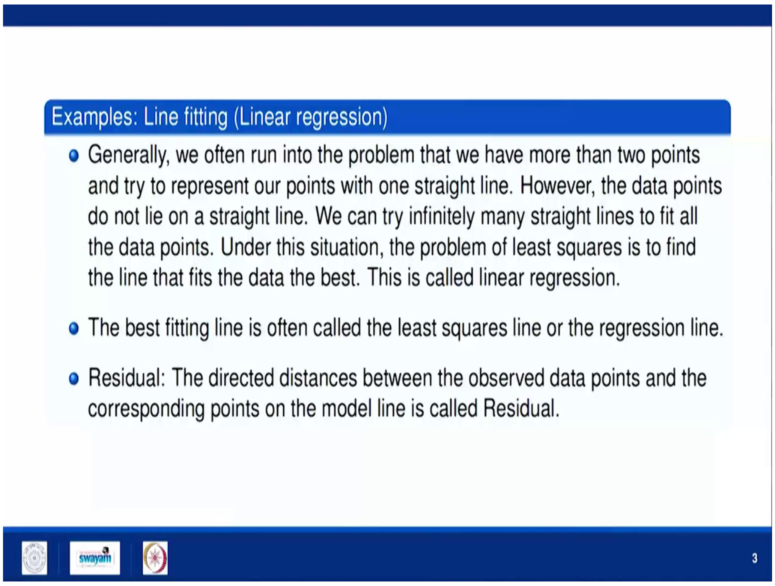
If we talk about underdetermined system then what we are having. So, in underdetermined system we are having, that is the less number of rows than the unknown variables. So, for example, we are having some system like this x 1 plus x 2 minus x 3 equals to 2 and 2 x 1 plus x 2 plus x 3 equals to 4.

So, here we are having two equations and n equals to 3, that is 3 unknown variables. So, in this case my coefficient matrix will become 1 1 minus 1, 2 1 1 that is my matrix A which is 2 by 3. And then I am having the column x, which is the column of unknown variables equals to 2 4 that is the right hand side vector b.

So, here x is 3 by 1 and it is 2 by 1. So, in this case if you can notice the coefficient matrix is a fact matrix that is we are having more number of columns. So, in this kind of situation we will be having always infinite number of solutions, because if you are having n unknown variables and m equation.

So, any n minus m variables can be chosen arbitrary and by choosing those for the rest of the system we can find out a solution. So, we can choose those n minus m variables in arbitrary in infinite many way. So, we will be having infinitely many solutions. But in this lecture I will talk about a special case of solution that is called minimum normed solution for underdetermined system.

(Refer Slide Time: 07:14)

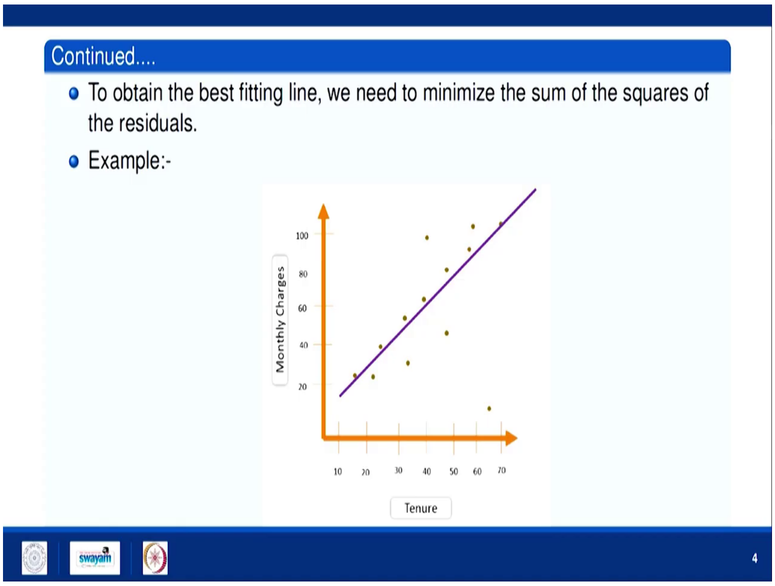


Let us come back to the example of over determined system. So, one of the very basic example of over determined system is linear regression, that is the line fitting. So, generally we often run into the problem that we have more than two points and try to represent our points with one straight line.

So, suppose we are given 10 points and we have to fit a line, which is the best fit line from these 10 points. However, these 10 data points which I am talking do not lie on a straight line. So, we can try infinitely many straight line to fit all the data points, under this situation the problem of least square is to find the line that fits the data the best.

Here best means which is having the minimum residual error, this is called linear regression. The best fitting line is open called the least square line or the regression line also. And based on that we say for over determined system the solution is least square approximation solution. As I told you here best means which is having the minimum residual error.

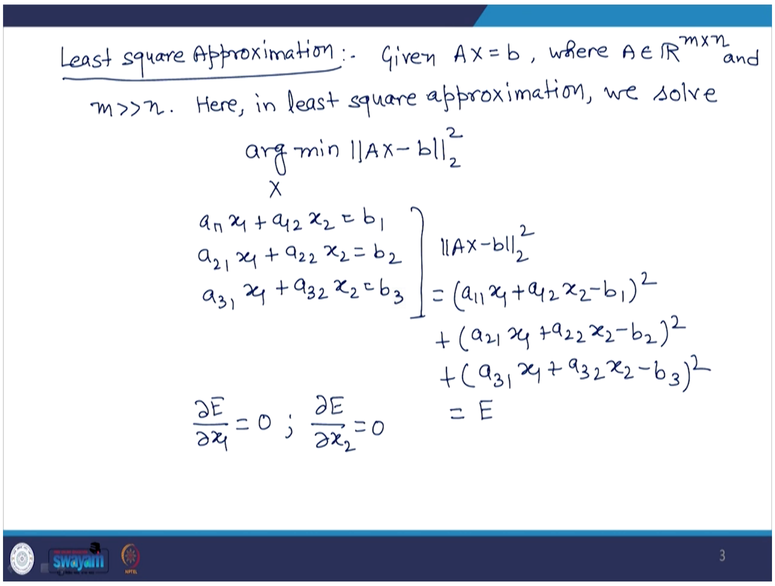
(Refer Slide Time: 08:47)



So, what is residual? The directed distance between the observed data points and the corresponding points on the model line is called the residual. And then sum of the square of all those residual is called the residual error. So, just see this example, here we are having this tenure that is in months and here monthly charge. And we are having in 2 dimensional plane these data points, so certain data points.

So, this is the best fit line in this data. So, what I have already told you, to obtain the best fitting line we need to minimize the sum of the square of the residuals as we are doing here. So, residual for this these are the residual, the perpendicular distance from line to all points. So, these are the residual and we have to minimize the sum of the square of all the residuals. How to do that?

(Refer Slide Time: 09:33)



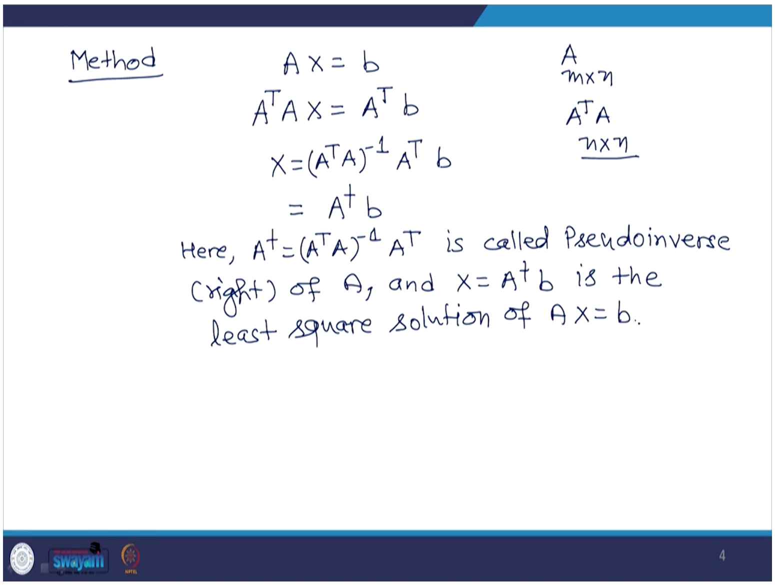
So, least square approximation. So, given A X equals to b, where A belongs to R m by n, that is A is m by n matrix having real m (Refer Time: 10:12). And m is quite bigger than n that we are having an over determined system. So, in this case, how to find out least square approximation?.

So, here in least square approximation, we solve the optimization problem that is the, we minimize the Euclidean norm between A X and b. So, X is a vector b is a vector and we minimize it, how? So, for example, if you are having let us say 2 by 3 system which is given as a 1 1 x 1 plus a 1 2 x 2 equals to b 1, a 2 1 x 1 plus a 2 2 x 2 equals to b 2 and a 3 1 x 1 plus a 3 2 x 2 equals to. So, here what is the meaning of A X minus b norm square.

So, this is nothing just a 1 1 x 1 minus a 1 2 x 2, sorry plus minus b 1 square plus, a 2 1 x 1 plus, a 2 2 x 2 minus b 2 square, plus a 3 1 x 1 plus, a 3 2 x 2 minus b 3 square. So, now this is my some of the residual errors. Now least square means we have to find out x 1 and x 1 which minimize this particular sum of the residuals. So, for minimizing this what I will be having, I have to put the necessary condition of the minima that is del e over del x 1 equals to 0 and del e over del x 2 equals to 0.

So, from this I will get two linear equations in x 1 and x 2 and by solving those two linear equations I will get the value of x 1 and x 2, which minimize the sum of the squares of the residual errors. So, how to do it?

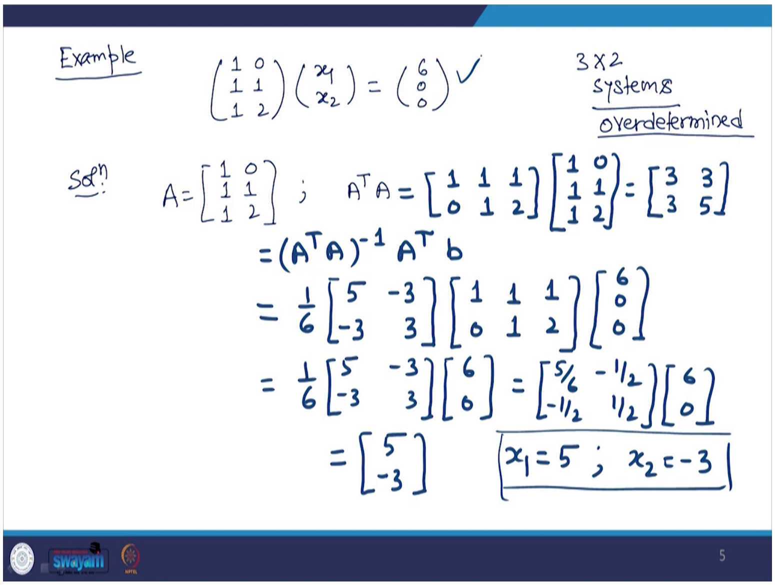
(Refer Slide Time: 13:25)



So, the easiest way is, we are having A X equals to b. What you do you multiply both side by a transpose. So, A transpose A X equals to A transpose b. So, here it a is m by n matrix then A transpose A will become n by n matrix. If the rank of A is n then what we are having?

We will be having x equals to A transpose A inverse, which is a square matrix and since rank is n. So, it will be a full rank matrix. So, inverse will exist into A transpose b. This I can write pseudo inverse of A into b. Here this equals to A transpose A inverse into A transpose is called pseudo inverse, that is the right pseudo inverse of A. And X equals to pseudo inverse of A into b is the least square solution A X equals to b.

(Refer Slide Time: 15:28)



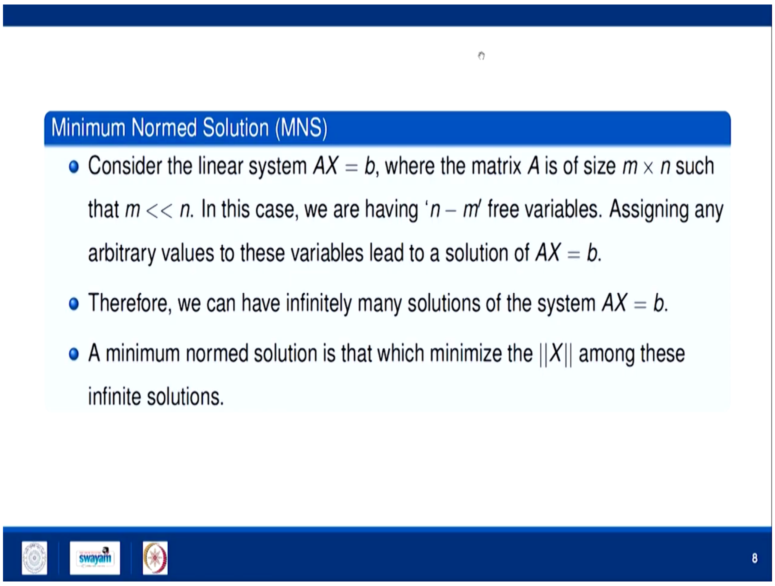
So, let us take example. So, what I am having 1 0 1 1 1 2, into x 1 x 2 equals to 6 0 0. So, it is a 3 by 2 system, means three equations in two unknowns x 1 and x 2. So, it is an over determined system.

So, we have to go for least square approximation. So, here A is 1 0 1 1 1 2. So, I will calculate A transpose A. So, A transpose A equals to. So, it will become 1 0 1 1 1 2 and then, A is 1 0 1 1 1 2. So, this comes out to be 3 and then 3 3 and then 5.

Here A transpose A inverse into A transpose b. So, what is A transpose A inverse it is 1 by 6. So, 6 is the determinant of A transpose A and then 5 then minus 3 minus 3 and 3. So, this is A transpose A inverse into A transpose. So, A transpose is 1 0 1 1 1 2 into b b is 6 0 0.

So, this will be 1 by 6, let us have it 5 minus 3 minus 3 3 and then this one will become 6 and then 0. Now it will be 5 by 6, minus 1 by 2, minus 1 by 1, 1 by 2 into 6 0, which will be 5 and then minus 3. So, least square approximation is x 1 equals to 5 and x 2 equals to minus 3. So, in that way we are able to find out least square solution of this over determined system.

(Refer Slide Time: 19:12)

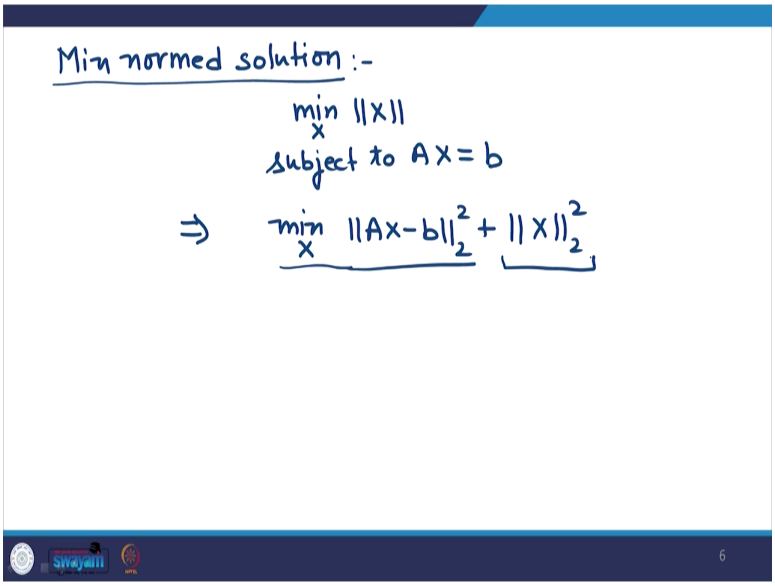


Now, let us discuss the another case that is the underdetermined one. So, consider the linear system A X equals to b, where the matrix A is of size m by n, such that m is very less than n. So, for example, you are having only 10 equations and anything more than 10 variables like 15, 20 or let us say 1000 variables. In this case we are having n minus m free variables assigning any arbitrary values to these free variables lead to a solution of A X equals to b.

Now, these n minus m variables we can assign arbitrary values in infinite way, hence this A X equals to b will be having infinite number of solutions. So, we can have infinitely many solutions of this system, a minimum norm solution is that which minimize the norm or that is the length of vector as X among these infinite solutions.

So, what we are interested? Out of those infinite numbers of solutions we are looking for a solution which is having the minimum norm and such a solution is called minimum normed solution minimum normed solution.

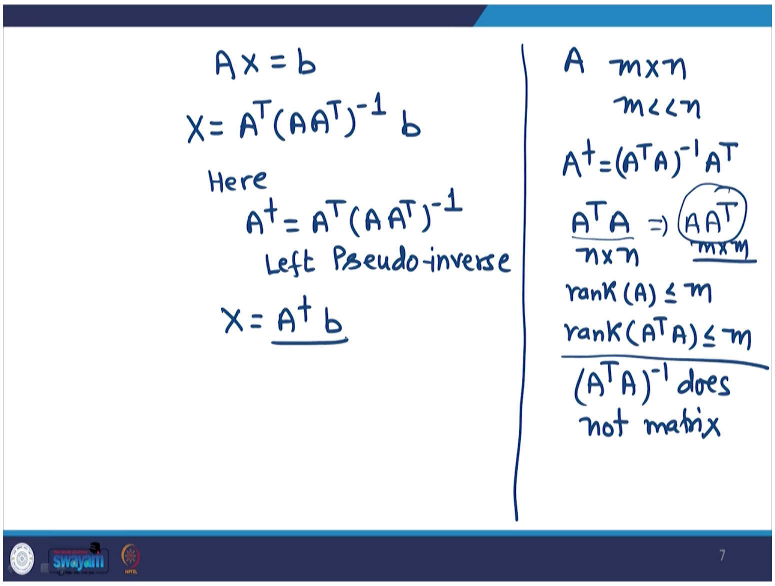
(Refer Slide Time: 20:34)



So, mathematically how can we pause this problem? So, we have to find out X, which minimize the norm of X, subject to A X equals to b. So, A X equals to b that it should be a solution of this linear system, this linear system will be having infinitely solution. So, out of those infinite solutions I am interested in the solution which is having the minimum norm.

So, how can I write it? I can write it minimize A X minus b. So, let us take two norm plus X. So, I have taken these two together. So, if you just compare with the earlier one least square approximation case, there I was having only this objective function but here I am having this minimum norm condition extra. So, how to solve such a system? Again we will use the concept of pseudo inverse.

(Refer Slide Time: 22:13)

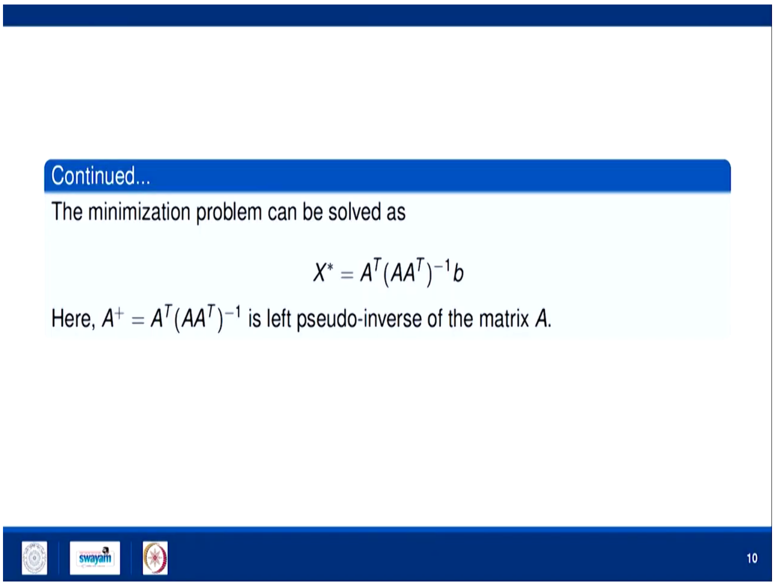


So, what I am having a system A X equals to b. Here A is m by n matrix and m is less than n. In the case of least square approximation you have seen that pseudo inverse is A transpose A inverse into A transpose.

Now if I calculate a transpose a here it will be of size n by n, while what is the rank of A? Rank of A is less than equals to m. So, if I assume even m. So, rank of A T A is less than equals to m. So, even though A transpose A is having rank m, but size is n by n, then n is bigger than m.

So, A transpose A inverse does not exist because it is rank deficient matrix having the determinant 0. So, here we cannot go like we have done in case of least square approximation. So, what is the solution? So, solution is simple.

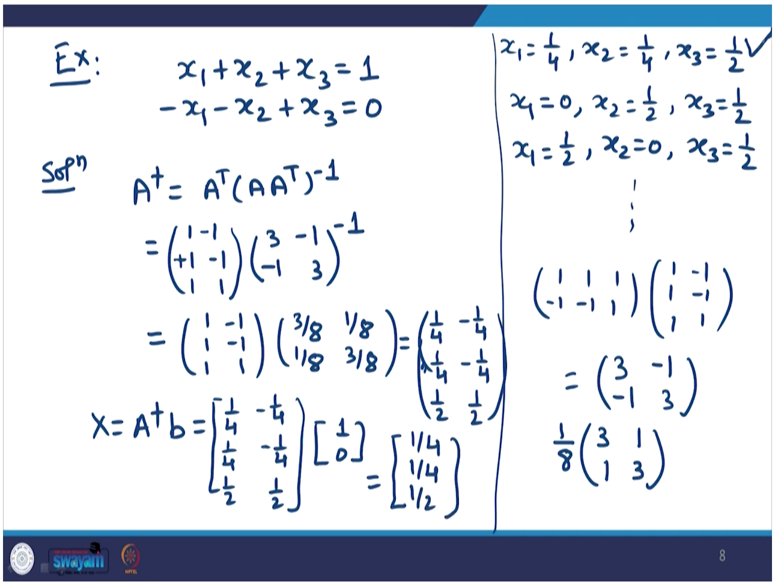
(Refer Slide Time: 23:40)



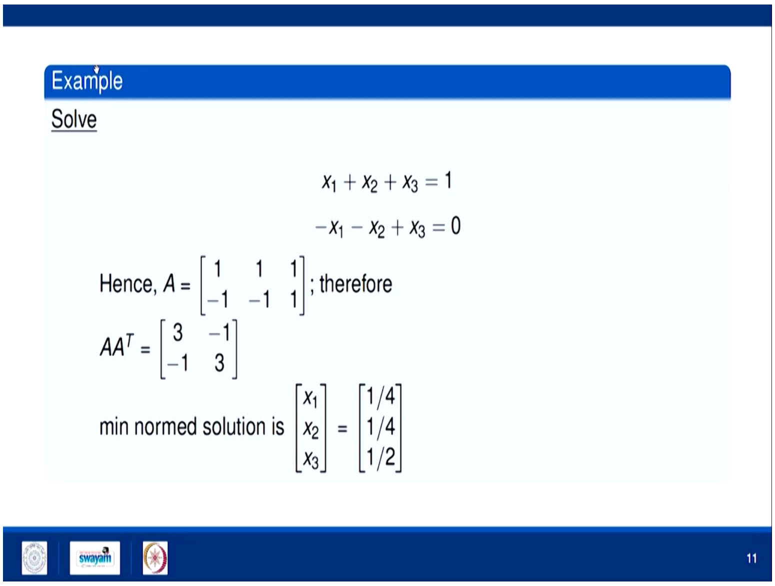
Here we will calculate the pseudo inverse is. So, instead of A transpose A, I will go for A A transpose, which is again m by m matrix. And if the rank of A transpose A A T is m then what I will be having, I will be having this as a full rank matrix so inverse exists and I will use this concept.

So, here pseudo inverse of A is, A transpose A, A transpose inverse this is called left pseudo inverse. And the solution X equals to the left pseudo inverse of A into b, will give you the minimum normed solution of the system A X equals to b which is an underdetermined system.

(Refer Slide Time: 25:00)



(Refer Slide Time: 25:04)



So, let us take an example of this. So, example is x 1 plus x 2 plus x 3 equals to 1. And another equation is minus x 1 minus x 2 plus x 3 equals to 0. So, it is an underdetermined system because we are having three unknown variables x 1, x 2 and x 3 and only two equations. So, this system will be having infinitely much solution.

For example, you can have one of the solution as like, x 1 equals to 1 by 4, x 2 equals to 1 by 4 and x 3 equals to 1 by 2. So, this will satisfy both the equations another solutions you can have something like, x 1 equals to 0, x 2 equals to half and x 3 equals to half. Similarly one of the solution may be x 1 equals to half, x 2 equals to 0, x 3 equals to half.

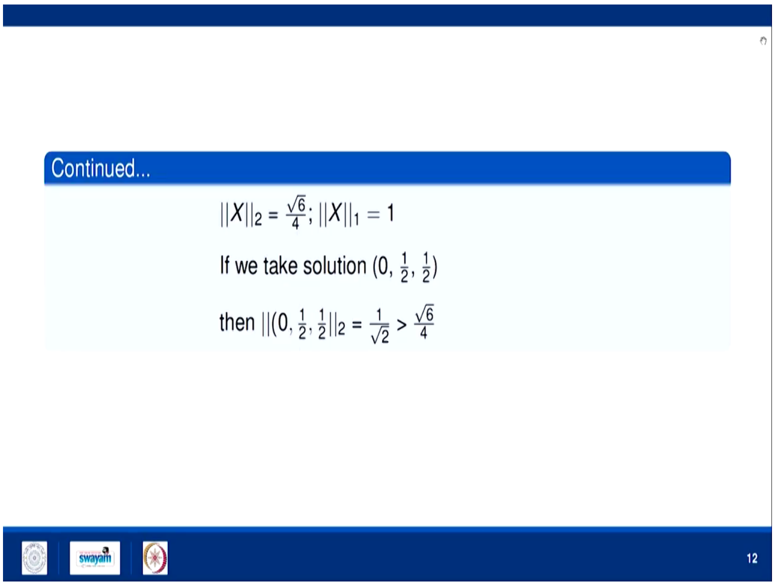
So, in that way you will be having infinitely many solutions. These are some examples I have just directly calculated. But out of all those infinite solution I am interested in a solution which is having the minimum norm. So, again we will use whatever we have learned. I will calculate pseudo inverse of A, that is A transpose, A into A transpose inverse this one. So, here A transpose will become 1 1 minus 1 minus 1.

Sorry 1 minus 1, 1 minus 1 and then 1 1. This is A transpose A into A transpose will become 1 minus 1, 1 minus 1, 1 1 into 1 1 1 minus 1 minus 1 1. So, this comes out to be 3 minus 1 minus 1 and 3. So, 3 minus 1, minus 1 3 inverse. So, this will be. So, inverse of this will be 1 upon 8 and then 3 1 1 3.

So, it will be 1 1 1 minus 1 minus 1 1 and then 3 by 8 1 by 8, 1 by 8, 3 by 8. So, this equals to 1 by 4 and then 1 by 8 minus 3. So, minus 1 by 4, again 1 by 4 minus 1 by 4 and then the last row will be 1 by 2, 1 by 2. Now minimum normed solution is; this into b.

So, it is 1 by 4, 1 by 4, 1 by 2 minus 1 by 4, minus 1 by 4, 1 by 2 into b and b is 1 0. So, this b got 1 by 4, 1 by 4 and 1 by 2. So, this solution is the minimum normed solution just to verify it.

(Refer Slide Time: 29:18)

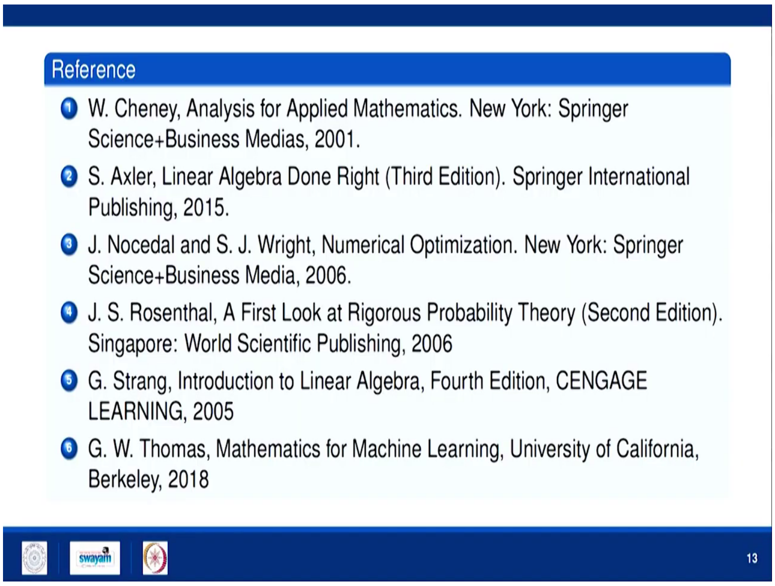


Here length of this solution is root 6 by 4 and n 1 norm is 1. If we take another solution, let us say 0, half, half then it is Euclidean norm is 1 by root 2 which is of course, bigger than square root 6 upon 4.

Similarly, we can verify for other solutions. So, this solution which I have taken here 1 by 4, 1 by 4, 1 by 2 will be having the minimum length, that is which is root 6 upon 4. Any other solution because it will be having infinite number of solutions we will be having normed bigger than root 6 by 4.

So, in this lecture we have learned about least square approximation, which is a solution method for solving over determined system. And the same time in case of underdetermined system we have learned about minimum normed solutions.

(Refer Slide Time: 30:19)



These are the references for this lecture. In the next lecture we will learn a very beautiful concept of linear algebra, which is quite useful in machine learning and image processing that is singular value decomposition.

Thank you very much.