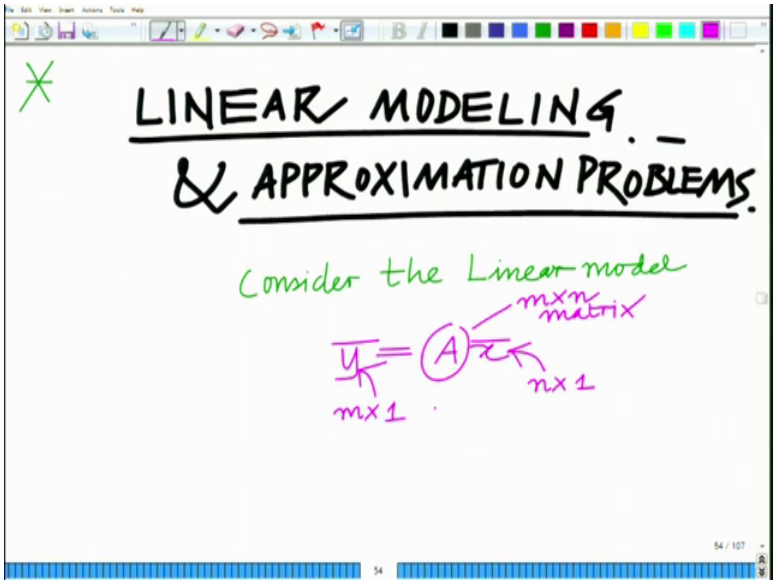
Applied Optimization for Wireless, Machine Learning, Big Data

Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 41
Linear modeling and Approximation Problems: Least Squares

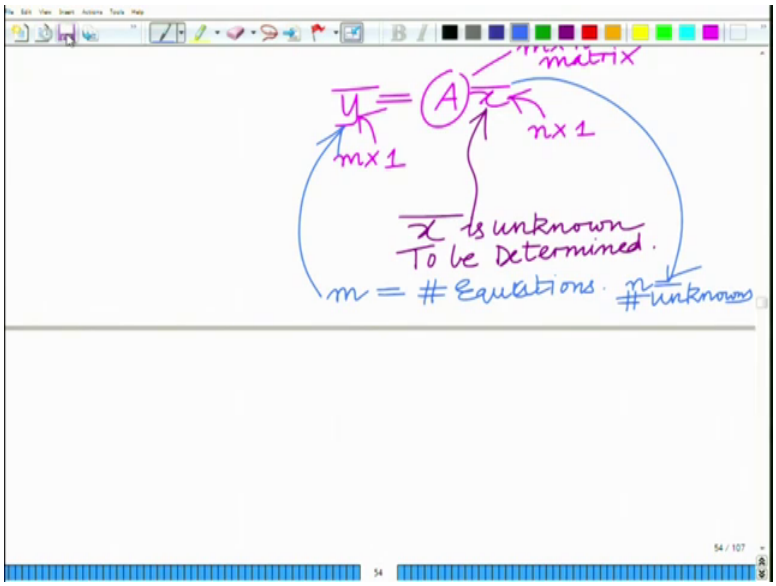
Hello, welcome to another module in this massive open online course. So, in this module, let us look at another class of problems or another class of optimization problems, specifically pertaining to Linear modeling and Approximation, which arise very frequently in various applications of course both engineering, science and so on so. This is a very important class of problems.

(Refer Slide Time: 00:38)



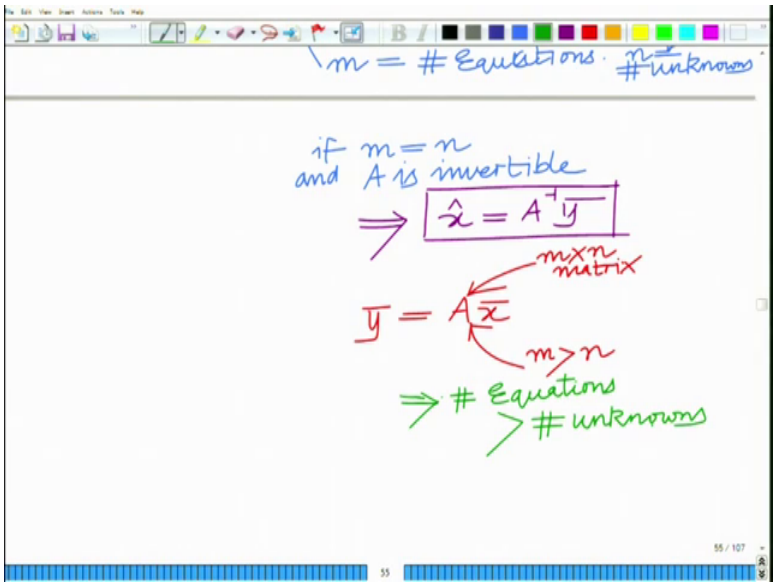
This is termed as linear wherever you have models, and typically more most of them or frequently they are modeled as linear model. So, we have linear modeling and approximation problems, linear modeling and approximation problems. And what you can see in this is that if you consider now consider the linear model, general linear model can be described as follows. Consider the linear model y bar equals A x bar, where we have A, this is in general an m cross n matrix, and which implies of course that if x is a vector, x is a n cross 1 vector. And y bar is a naturally an m cross 1 vector.

(Refer Slide Time: 02:18)



Now, what we assume or typically in this model, this vector x bar is unknown, which has to be determined. So, x bar is unknown and it has to be determined so, x bar is to be determined. And now frequently what you have also in such a system, now of course let us start by considering a simple example. Let us say A is square matrix that is m is equal to n ok. If m is a number of rows, n is a number of columns of A. Let us now if you look at this system, you can see what is m in this system, m is basically the number of equations, it is the dimension of y. So, m equals number of equations. And n equals number of unknowns ok.

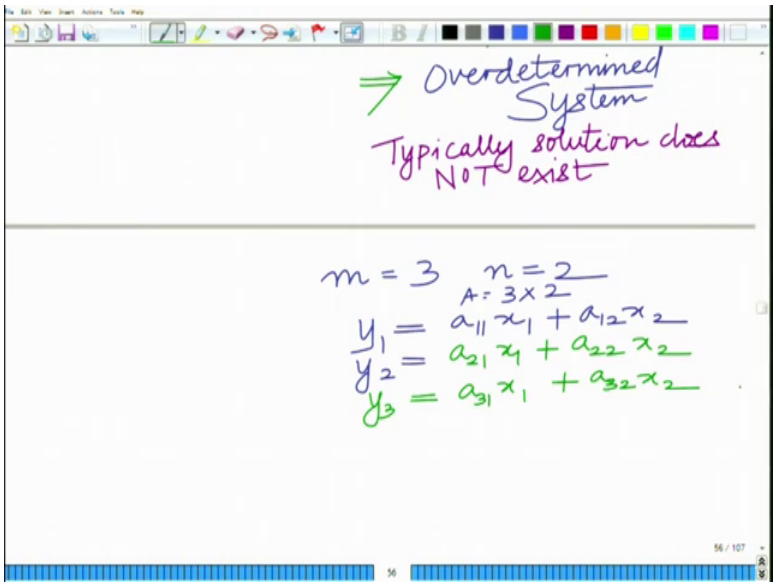
(Refer Slide Time: 03:31)



Now, what happens in this is that let us assume a simple scenario to begin with if m equals n, and this everyone would know m equals n. And A is invertible, it is if equals m and A is invertible, remember this is as to be given. Now, this implies I can find y y bar is equals to x bar, I can determine x hat or the estimate of x equals A inverse y bar ok. This I think is a typical solution for the linear system, which most people would know almost students would know.

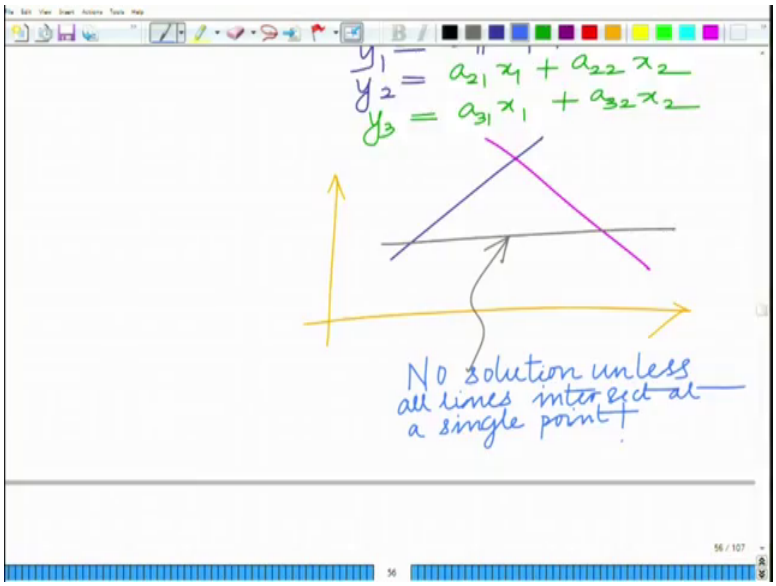
Now, however frequently what we have is that you have this y bar is well, it is not exactly equal to A x bar its. Actually, A x bar plus n when n is the noise, I write it as y bar equal this is the linear model, y bar equals A x bar. And A is your m cross n matrix, and m is greater than n, m is greater than n ok. So, what that means is this implies that number of equations is greater than number of unknowns.

(Refer Slide Time: 04:57)



And which also implies that this is basically an over determined system, because number of equations is greater than this is basically an over determined system. Now, of course over determined system frequently, this has no frequently no solution exists, unless the vector y bar belongs the column space of A. So, frequently typically because of the noise in this system, typically does not does not exist. For instance, you can take a simple example. Let us take m equal to 3, what we have is we have y 1 equals a 1 1. So, A equals 3 cross 2 matrix. So, we have y 1 equals a 1 1 x 1 plus a 1 2 x 2; y 2 equals a 2 1 x 1 plus a 2 2 x 2; y 3 equals a 3 1 x 1 plus a 3 2 x 2.

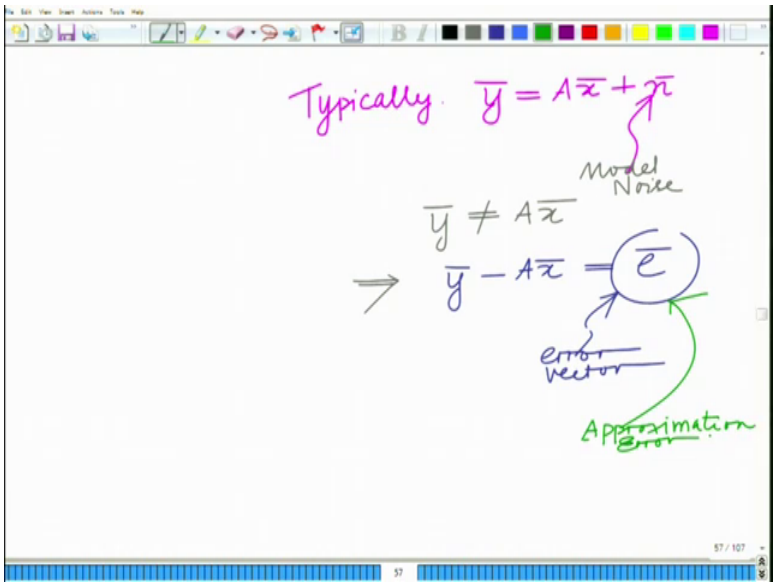
(Refer Slide Time: 06:31)



Now, you can see these represent three line. So, what happens when you plot this things, you have three lines. So, you basically have three lines each represents a line ok. Now, the solution now unless all the lines intersect no there is no solution unless all the lines intersect at a common point all right. So, you see this point. So, you have three lines that is three equations two unknowns ok. So, there is no solution unless all the lines intersect at a there is no solution unless all the lines intersect at a single point all right.

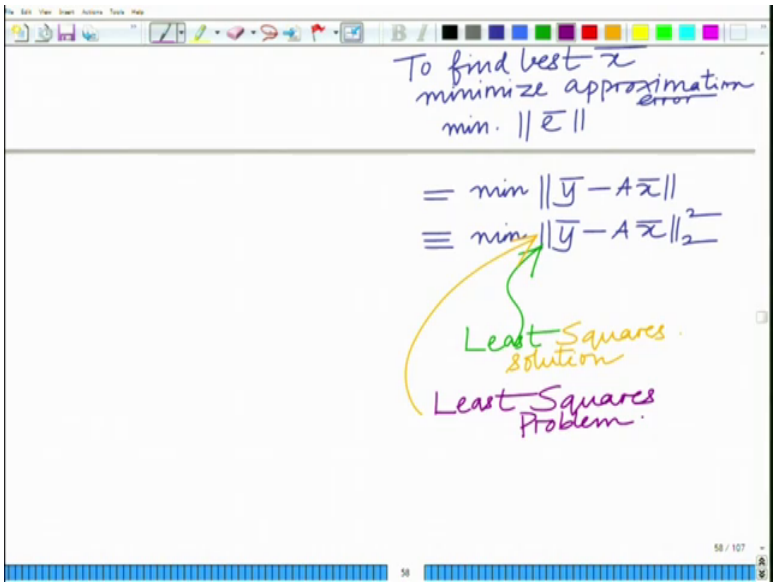
So, you have three equations basically, two unknowns. And frequently if you take three lines at random, then they will naturally not intersect. It is highly unlikely that they all intersect at a single point, which means there will be no solution. So, in such a scenario you will have to find an approximate solution or what I mean by that is some solution that best fits the model or best explains the observed vector y. This is also known as the maximum likelihood vector x ok, so it is interesting.

(Refer Slide Time: 08:00)



So, typically and this is because this happens because of the noise in the model. Typically, what happens is you have y bar equals A x bar plus n bar ok. So, this is what is this is your model noise. And what happens because of this model noise this means, your y bar not equals is not equal to A x bar, which means if you form y bar minus A x bar, there for any x bar there is always, because there is no solution for any x bar, there is always an error. This is termed this is basically your model error or your approximation error vector. So, you can also call this as a error vector, which is basically the approximation error. This is basically the approximation error.

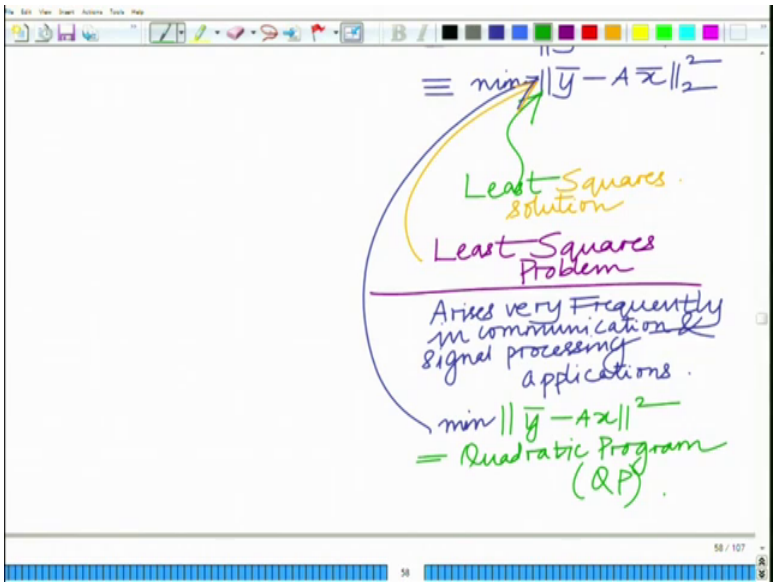
(Refer Slide Time: 09:11)



And we have to find x bar, you have to find best x bar that is which best explains y bar, what do you do is minimize the approximation error. Now, error is of course a vector. What does it mean to minimize the error, we will simply minimize the norm of this error vector. So, we will minimize the norm of the error vector, which is basically equal to minimizing norm of y bar minus A x bar, which is basically equivalent to minimizing norm square of y bar minus A x bar square ok.

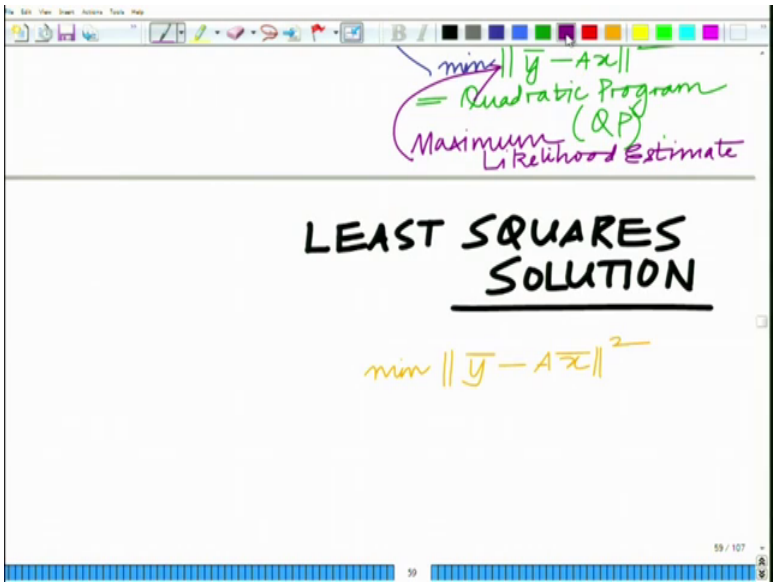
And this problem where you are and this is the two norm square, the l two norm. And when you are minimizing the norm square, so this is the vector x bar, which gives you the least squared error norm. So, this is known as the least square solution this is known as the least square solution or this is known as the least squares problem, in fact this is known as the least squares problem.

(Refer Slide Time: 10:55)



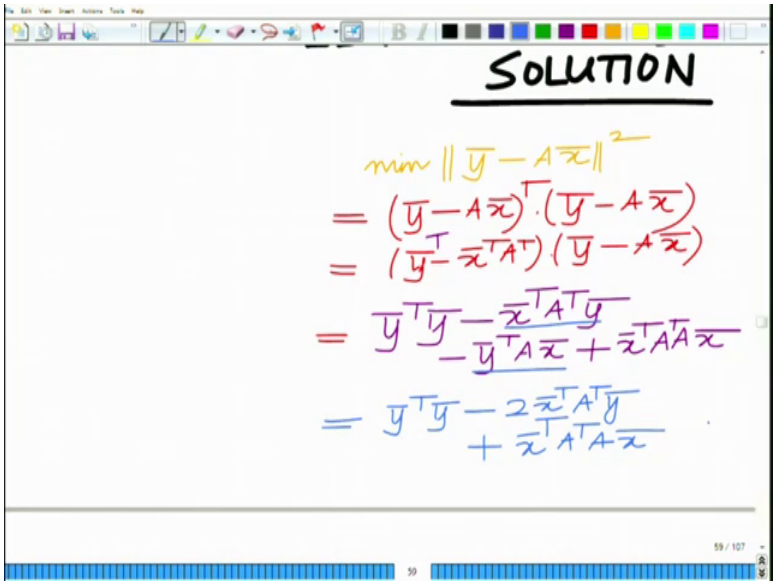
And this is very popular in communication. This is arises, when we are going to express the show some example, so this arises very frequently in communication and signal processing applications ok. So, this arises very frequently in communication and (Refer Time: 11:39). And in fact if you look at, this is nothing but a this is a quadratic objective function or this also termed as a quadratic program. So, this minimize norm of y bar minus A x bar square. This is termed as a it is a quadratic objective function. It is termed as a quadratic program or basically a QP ok. And finding the solution of this QP gives the best estimate this is also known as the maximum likelihood estimate. So, the solution of this gives the least squares problem, gives the maximum likelihood estimate. In fact, strictly speaking the maximum likelihood estimate in Gaussian noise ok.

(Refer Slide Time: 12:28)



So, just gives the what is termed as a or basically if one asks the question what is the vector x bar, which has the maximum likelihood right, which best explains the y, which has the maximum likelihood of having occurred that is the vector x bar, which basically minimizes the least squares, this which is the solution to the least squares problem that is it minimizes the square or it gives the least squared error least squared norm of the error vector ok. And this can be solved as follows and it is not very complicated. So, what we do is that to find the solution. So, we want to find the least square solution, and that can be found as follows. We want to minimize norm y bar minus A x bar square.

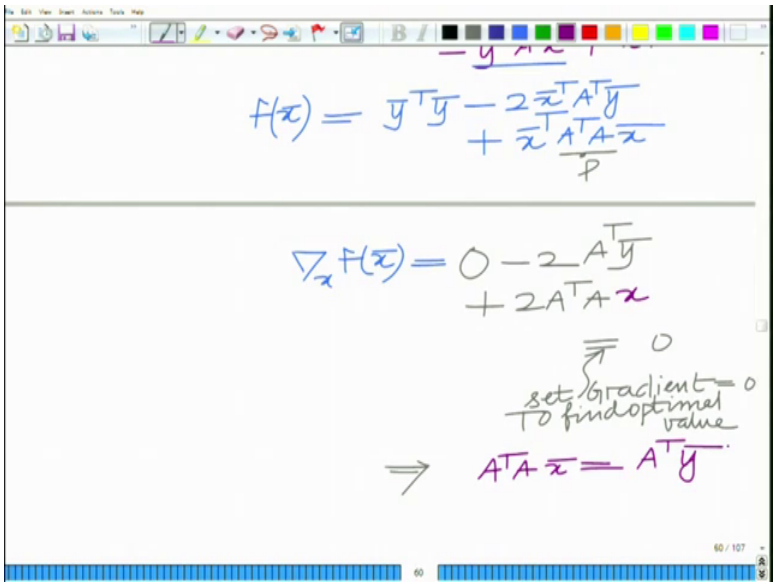
(Refer Slide Time: 14:12)



So, what we do is basically remember norm square of a vector, this is nothing but vector transpose times itself. So, this is y minus A x bar transpose times y bar minus A x bar, which is equal to y bar minus x bar transpose A transpose into y bar minus A x bar, which is equal to y bar transpose y bar minus x bar transpose A transpose y bar minus, now this is to be transpose minus y bar transpose A x bar plus x bar transpose A transpose A into x bar.

Now, if you can look at this these two quantities are the same, these are the transpose of each other and scalar quantities x bar transpose A transpose y bar y bar transpose A x bar. So, you can take the twice of one of these, so this will be finally simplified as y bar transpose y bar minus twice x bar transpose A x bar transpose A y bar plus x bar transpose A transpose A x bar.

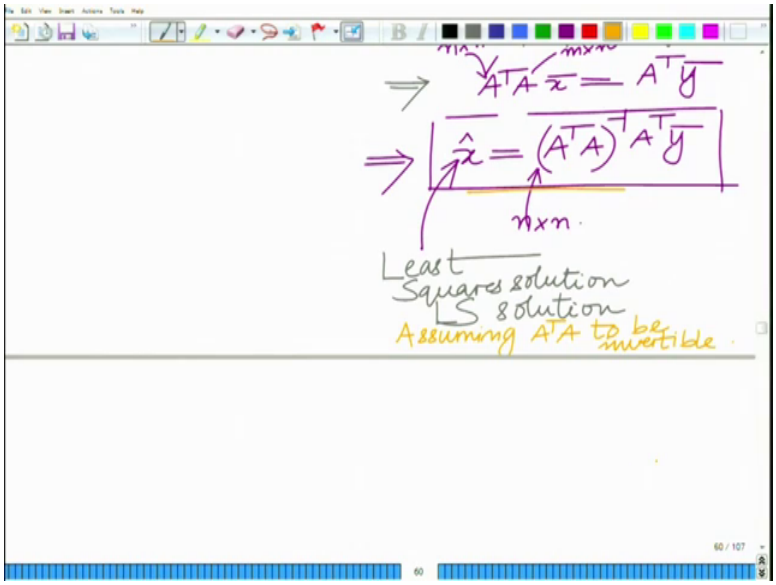
(Refer Slide Time: 15:42)



And now if you call this as now this is an objective function, because here you have no constraint, you have only the objective function as x bar. So, you take the gradient of F with respect to x bar. And now, we can see y bar transpose y bar gradient of that with respect to x is 0 minus twice; x bar transpose A transpose you can treat this as x bar transpose c bar, so the gradient is simply c bar. So, minus twice A transpose y bar plus x bar transpose A transpose x bar that is you can treat this as x bar; you can treat this as matrix P, it is positive semi definite, x bar transpose P x bar. So, the gradient is twice P x bar or twice A transpose A y bar.

Now, you said this equal to 0, to find the optimal value. So, you set gradient equal to 0 to find optimal value. This implies now if you said this equal to 0, what you now this 2's cancel ok, you cancel the 2. So, what you get is A transpose A y bar equals A I am sorry A transpose A, this has to be A transpose x bar, so we have A transpose A x bar equals A transpose y bar.

(Refer Slide Time: 17:16)



And what this implies is basically you have x bar or you can call it x hat, typically it used in the context of estimation. The optimal value of x is x hat equal to A transpose A inverse into A transpose y. Assuming A transpose A to be invert, because we can see A transpose A will always be a square matrix correct, because A is m cross n, and A transpose is n cross m. So, A transpose A will be n cross n matrix I am sorry, this is a transpose A will be n cross n. A transpose A inverse will also be n cross n. So, this is the least square solution.

Assuming, A transpose A is invertible, it is a very compact and elegant form. Least squares problem is also termed as the L S. And this assumes that A transpose A is invertible. This assumes A transpose A to be invertible, so that basically gives us the least square solution. And like this is one of the most fundamental problems in signal processing, and also for that matter in estimation, and communication and so far as so frequently.

A solution is very well known, and it is thought in a lot of courses. And in fact this forms this analysis of this problem from one of the staples of several course alright. And I think, this is one of the most important optimization problems with various applications that we are going to encounter in this course. So, we will stop here, and continue in other modules.

Thank you very much.