Applied Optimization for Wireless, Machine Learning, Big Data

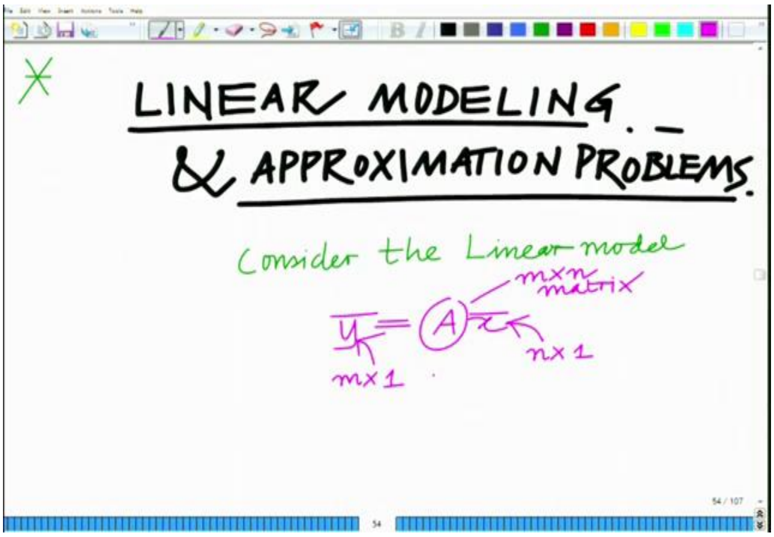
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 41
Linear modeling and Approximation Problems: Least Squares

Keywords: *Linear modeling, Approximation Problems, Least Squares Solution*

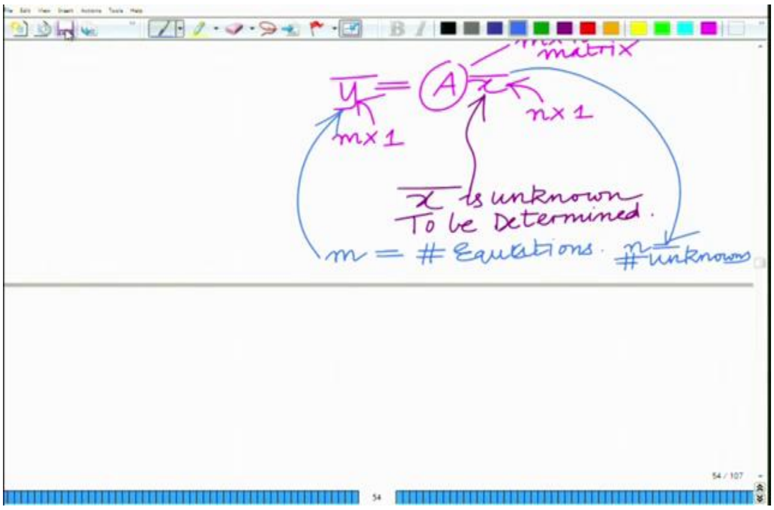
Hello, welcome to another module in this massive open online course. So in this module, let us look at another class of optimization problems, specifically pertaining to Linear modeling and Approximation, which arise very frequently in various applications across engineering, science and so on.

(Refer Slide Time: 00:38)



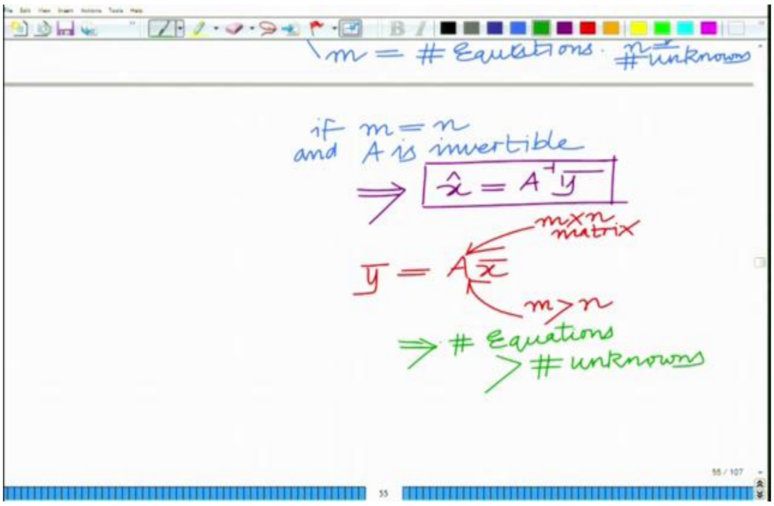
So consider a general linear model $\bar{y} = A\bar{x}$, where A is an $m \times n$ matrix, $\bar{x}$ is an $n \times 1$ vector and $\bar{y}$ is an $m \times 1$ vector.

(Refer Slide Time: 02:18)



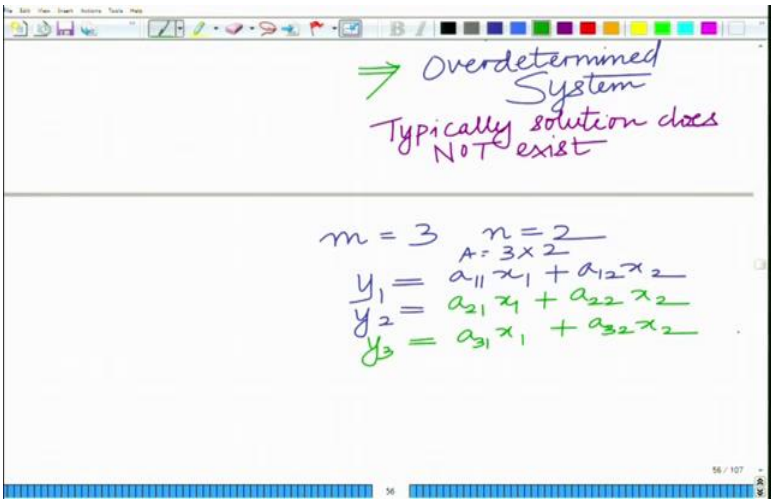
Now we assume this vector $\bar{x}$ is unknown and let us say A is a square matrix with m number of rows and n number of columns. If you look at this system, m is basically the number of equations and n is the number of unknowns.

(Refer Slide Time: 03:31)



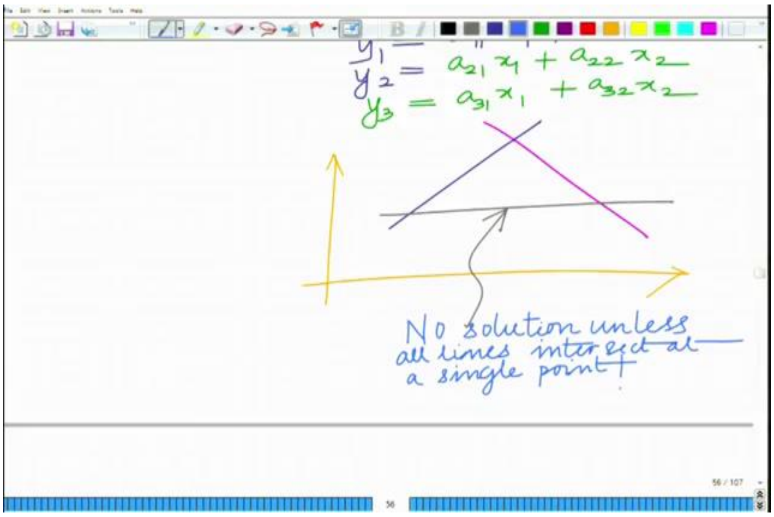
Now if m = n and A is invertible, this implies $x = A^{-1}\bar{y}$. This is a typical solution for the linear system. Now $\bar{y} = A\bar{x} + n$ where n is the noise and A is an $m \times n$ matrix and m > n which implies that number of equations is greater than number of unknowns.

(Refer Slide Time: 04:57)



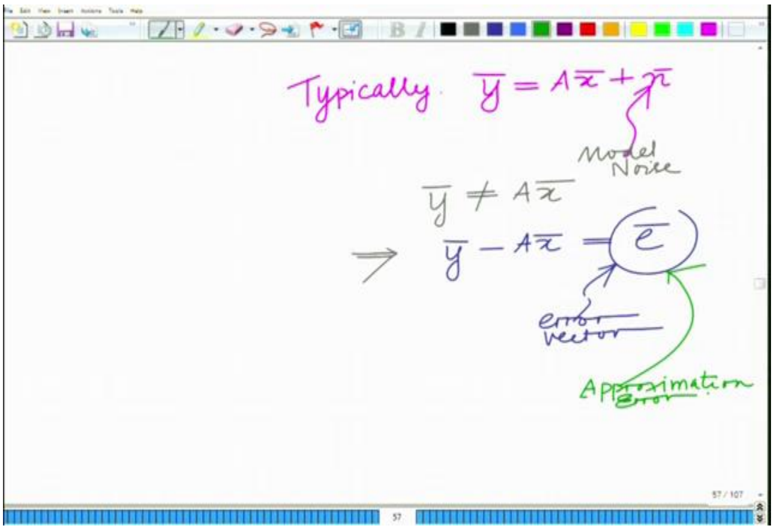
This also implies that this is an over determined system. Now this has no solution unless the vector $\bar{y}$ belongs the column space of A. For instance, you can take a simple example. Let us take m = 3 and n = 2 and we have $y_1 = a_{11}x_1 + a_{12}x_2, y_2 = a_{21}x_1 + a_{22}x_2, y_3 = a_{31}x_1 + a_{32}x_2$.

(Refer Slide Time: 06:31)



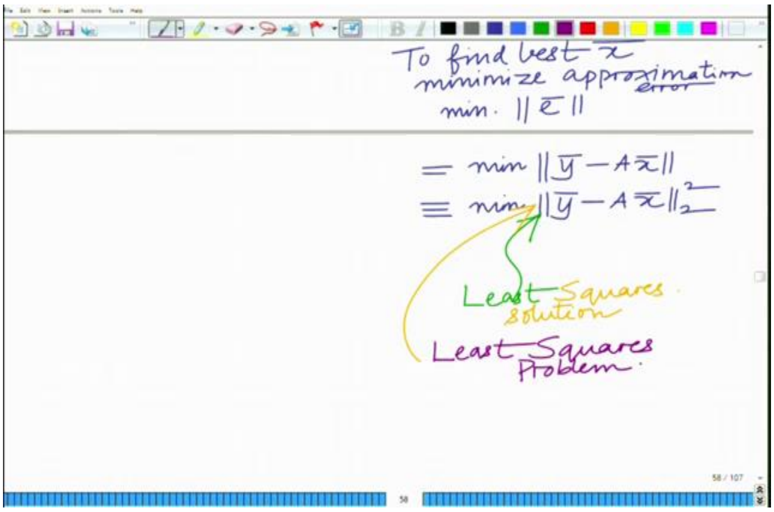
Now, you can see that these represent three lines and there is no solution unless all the lines intersect at a common point. So you have three equations and two unknowns and in such a scenario you will have to find an approximate solution that best fits the model. This is also known as the maximum likelihood.

(Refer Slide Time: 08:00)



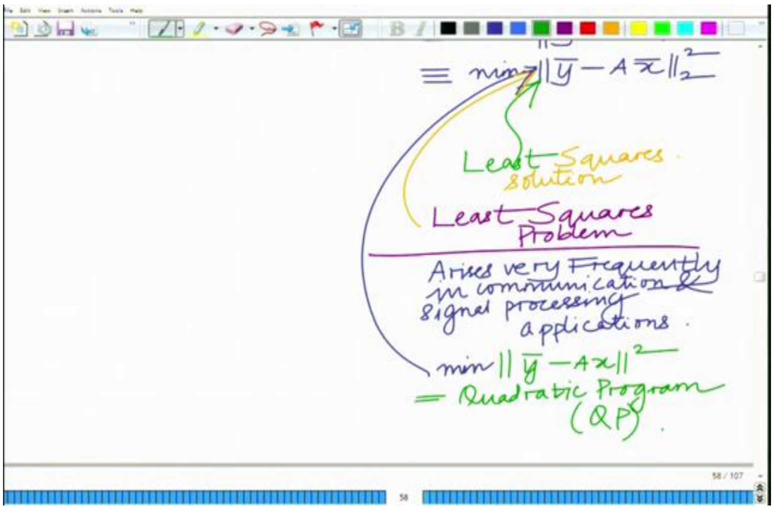
So because of this model noise $\bar{y} \neq A\bar{x}$ which means $\bar{y} - A\bar{x} = \bar{e}$ and this is an error. This is termed as model error or approximation error vector.

(Refer Slide Time: 09:11)



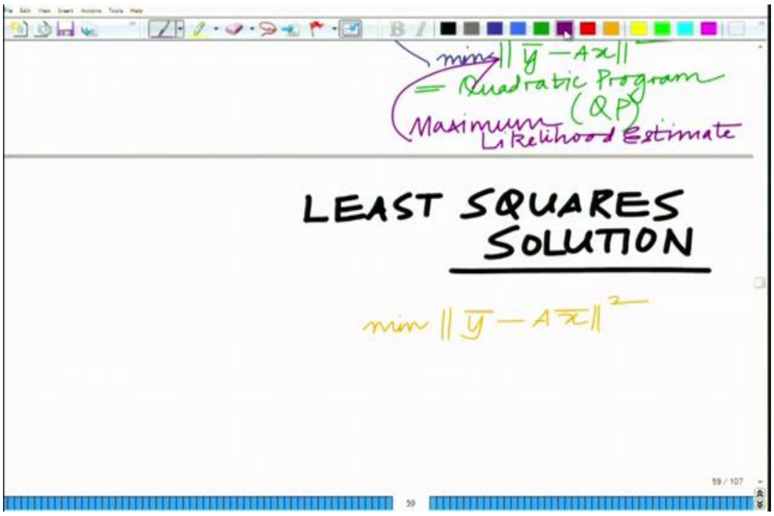
And we have to find the best $\bar{x}$ which minimizes the approximation error. Now this means that we simply minimize the norm of this error vector. So we have $\min\left\|\bar{y} - A\bar{x}\right\|_2^2$ and this is the l_2 norm. So this is known as the least square problem.

(Refer Slide Time: 10:55)



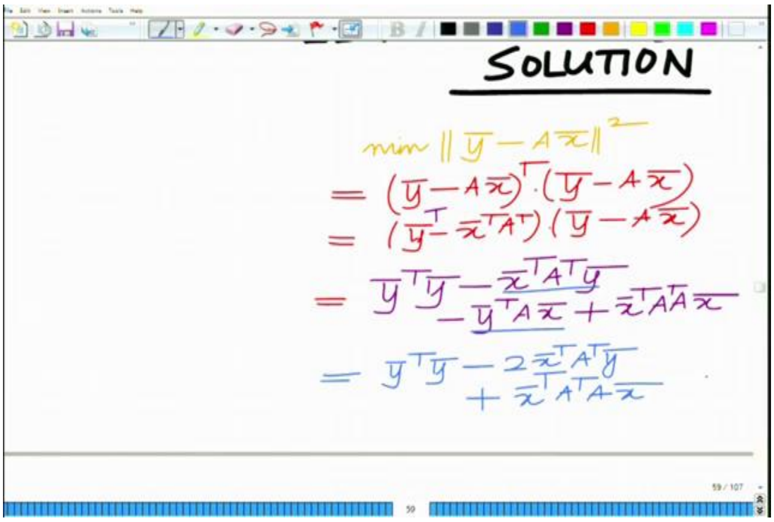
And this arises very frequently in communication and signal processing applications. In fact this is nothing but a quadratic objective function or this also termed as a quadratic program, a QP and finding the solution of this QP gives the best estimate, this is also known as the maximum likelihood estimate.

(Refer Slide Time: 12:28)



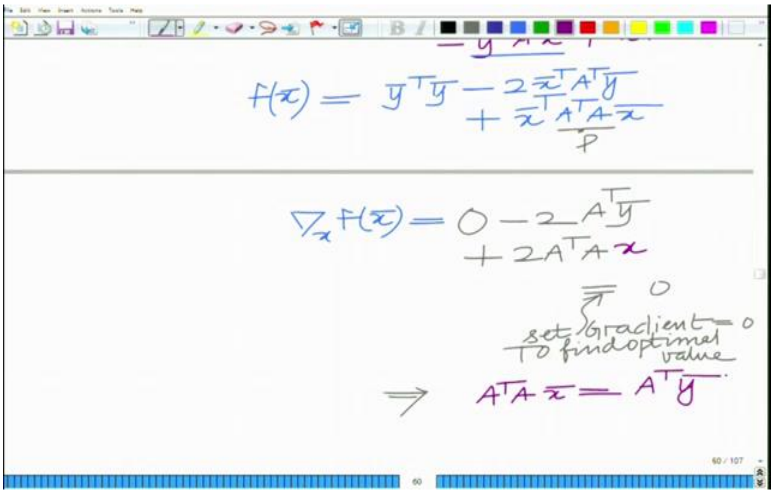
So we want to find the least square solution and that can be found as follows.

(Refer Slide Time: 14:12)



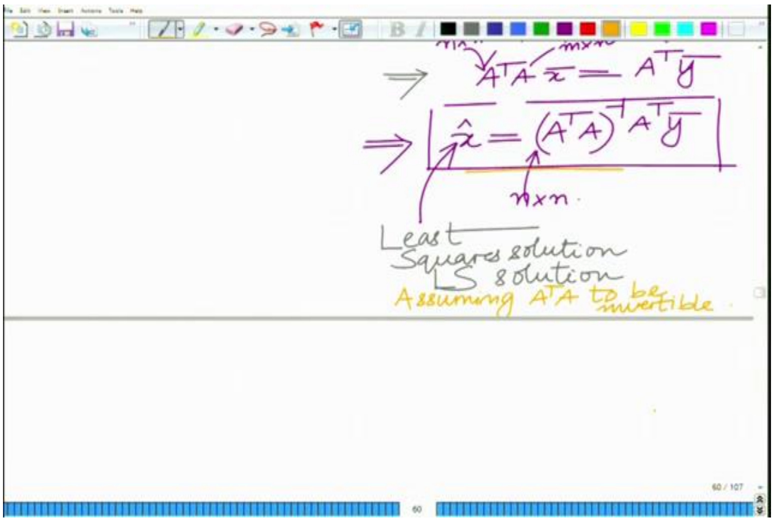
So we have $\min\left\|\bar{y} - A\bar{x}\right\|^2 = \left(\bar{y} - A\bar{x}\right)^T \left(\bar{y} - A\bar{x}\right) = \left(\bar{y}^T - \bar{x}^T A^T\right)\left(\bar{y} - A\bar{x}\right) = \bar{y}^T \bar{y} - \bar{x}^T A^T \bar{y} - \bar{y}^T A\bar{x} + \bar{x}^T A^T A\bar{x}$ so this is $f(\bar{x}) = \bar{y}^T \bar{y} - 2\bar{x}^T A^T \bar{y} + \bar{x}^T A^T A\bar{x}$.

(Refer Slide Time: 15:42)



And now if you call this as an objective function, we can take its gradient with respect to $\bar{x}$. So this will be $\nabla_x f(\bar{x}) = 0 - 2A^T \bar{y} + 2A^T A\bar{x}$.

(Refer Slide Time: 17:16)



The optimal value is $x = \left(A^T A\right)^{-1} A^T \bar{y}$ assuming A to be invertible. So this is the least square solution. So we will stop here and continue in other modules. Thank you very much.