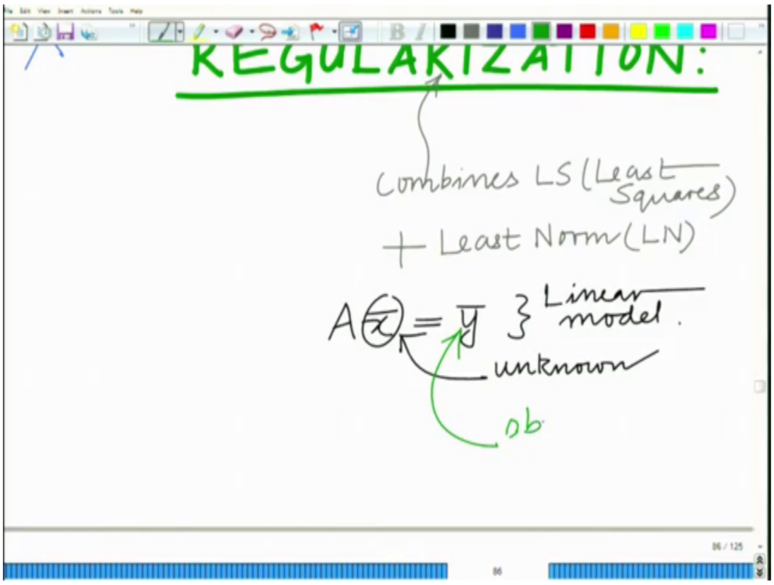
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 46
Regularization: Least Squares + Least Norm

Hello welcome to another module, in this massive open online course. So, we are looking at various convex optimisation problems, which are the Least squares and Least norm. To round to round up this discussion, let us look at another problem, which is essentially of combination of both these problem, that is the least squares and the least norm problem, which is termed as regularization.

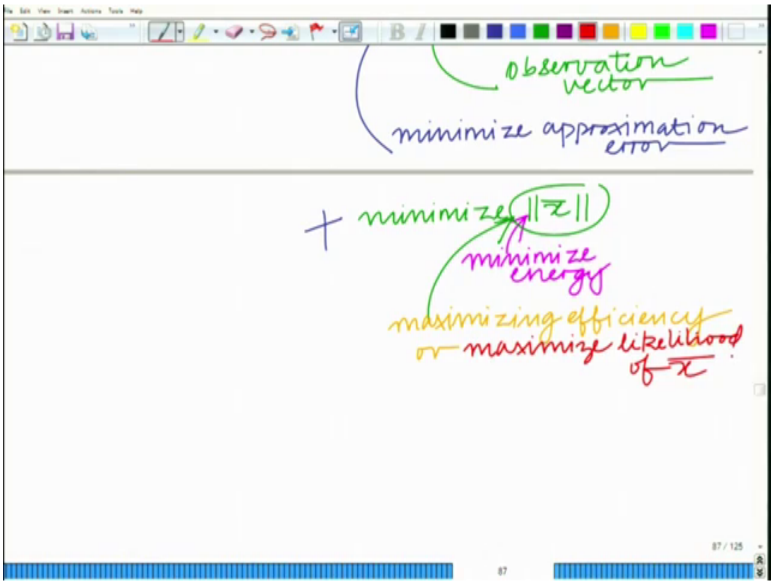
(Refer Slide Time: 00:37)



Or you can also think of this as a generalization of the least squares and the least norm of frame, (Refer Time: 00:50) it is termed as regularization ok.

And as I already said, what regularization does? Is that it combines, the least squares and least norm combines the L S that is your Least Squares plus the Least Norm frameworks. So, this is your L N and what did essentially does as follows remember, we said you we use the least squares framework, when we have a linear model or an over determined linear model and therefore, now consider this linear model A x bar equals y bar let us can be over determined or under determined, I will come to this in a moment.

(Refer Slide Time: 02:23)



So, this is A x bar equals to y bar this is your linear model and this vector x bar, this is an unknown vector similar to what we had previously, this vector y bar this is your observation matrix and the matrix is assumed to be known that is a matrix that is pertaining to your model or your linear model ok.

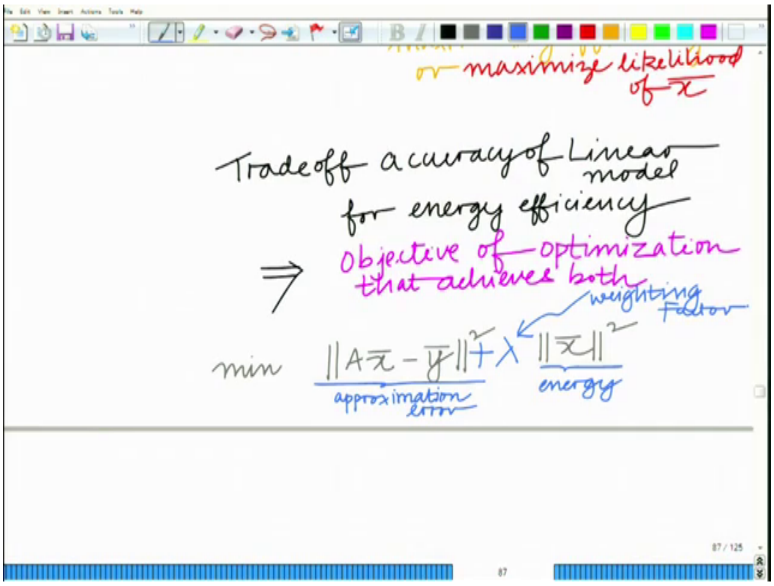
So, this is your observation vector, similar to the channel estimation problem you can think of this as the pilot vector that is the output vector that is a vector of observations, observed samples of the received samples at the receiver in the wireless system ok.

So now, we have a linear model in addition, we also desire. So, we want to minimise. So, we want to minimise the model error or minimise. So, it is desirable to minimise, the approximation error at the same time, we would also like to minimise plus, we would also like to minimise. Norm of x bar or basically we would also like to minimise the energy, that is minimise the energy, this can be thought of in various ways, that is basically to guaranty an energy efficient solution or it is known that vectors, which have lower energy have a higher probability. So, you are trying to maximize. So we considering, the prior probability of such factors arising in the problems.

So, we also trying to take factor in take into account that factor alright. So, this can be thought of as basically that, we can basically think of this as either maximizing the efficiency of the solution or you can also think of this as the fact that, you are trying to maximize the probability of x bar, this is a natural systems vectors x bar, which have lower energy have a higher probability. So, you have typically that is the scenario.

So, we would like to maximize the probability of x bar or you can also think of this is not probability, but rather the likelihood of x bar, I think that is a better word maximize the likelihood and therefore, we would want to have an trade off between the accuracy of the linear model and this probability slash energy efficiency ok.

(Refer Slide Time: 05:16)

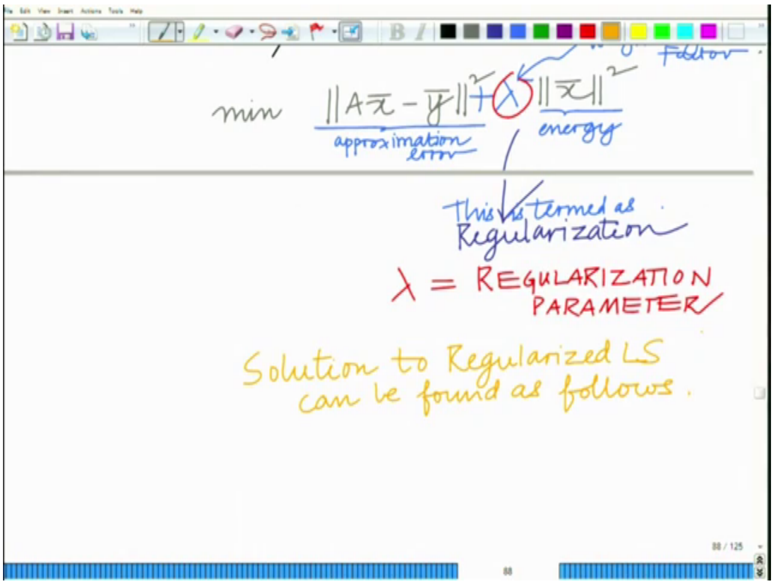


So, we would like to so, in summary, what we would like to do is we would like to it is not in either or we would like to do both this trade off accuracy of linear model for energy efficiency let us say. Implies we want to have an objective function, which reflects both correct we want to have an objective function or an optimisation, we want to have an objective of optimisation that achieves both that achieve both the above objectives and therefore, one can formulate the following optimisation problem, that is minimise remember previously to basically obtain the solution x bar, which best explains y bar or basically which minimises the approximation error, you have minimise the norm of A x bar minus y bar whole square, that is the least squares problem this minimises the model error.

Now, to maximize the at get a energy efficient solution, we would like to minimise the norm of x bar this minimise norm of x bar square, what we do now is the combine both of these alright. So, this basically is the approximation error and this is the energy of the solution and now you are minimising a linear combination of this. So, this is a weighting factor, now lambda here is not the Lagrange multiplier, this is a weighting factor or which can be a weight and therefore, you have a weighted objective. So, you have a weighted combination of these 2 objectives.

One is approximation error, one is a energy and you are minimising a weighted combination that is weighting the approximation error, weighing it by this energy alright, this process is known as regularization. So, you not just minimising the approximation error, but you are regularising this objective function by the addition of the energy by the addition of a component that is proportional to the energy of the solutions so this process.

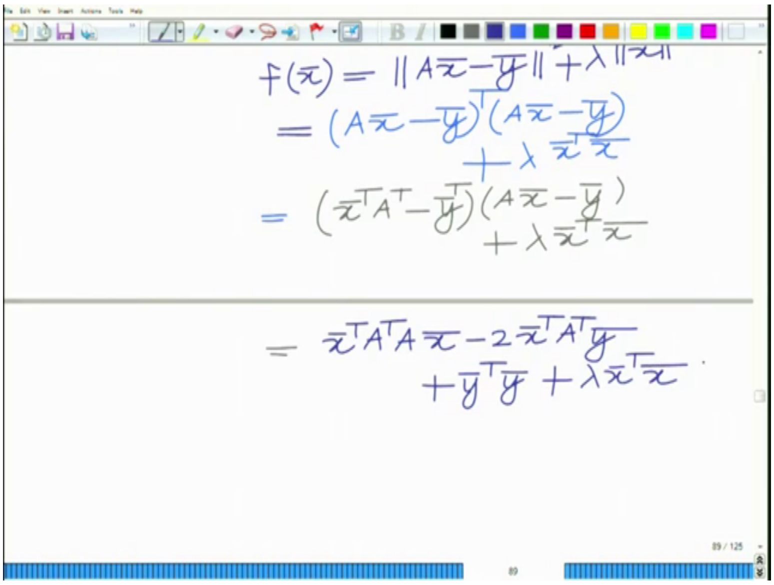
(Refer Slide Time: 08:04)



So, this basically encourages solutions that have lower energy. So, this constrains this process is termed as regularization and this. So, this process is termed as and this factor lambda, this lambda is termed as the regularization parameter.

This factor lambda is termed as regularization parameter and let us say, this is termed as, basically in a scenario where we would like to achieve both or it achieved of between the accuracy and as well as an energy efficient solution, one can apply this approach and the process the procedure to solve this is similar to that it similar to what we have seen before that is a expand, this objective function and it is rather once, you understood the previous paradigms, which one is going to be rather simple. So, the solution is formed as follows solution to the optimisation problem about.

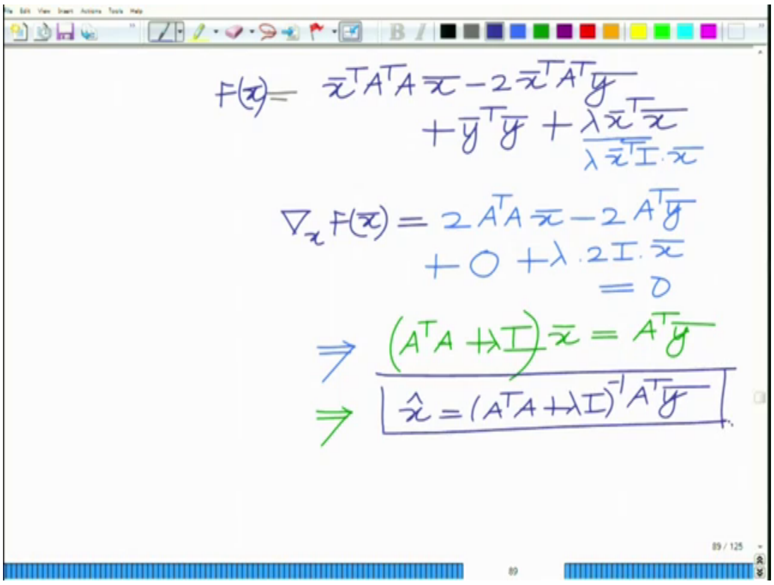
(Refer Slide Time: 09:52)



You can also think as this as the regularized least squares ok, the solution to the regularized L S can be found as follows well we have the objective function f of x bar this is norm A x bar minus y bar whole square plus lambda times norm x bar whole square, which I can write as norm square of a vector is vector times pose transition, that is for a real vector minus y bar transpose A x bar minus y bar plus lambda x bar transpose x bar which is equal to we can write this again, as again similar to what we have seen.

So, many times before x bar transpose A transpose minus y bar transpose, A x bar minus y bar plus lambda x bar transpose x bar, this is equal to well i can write this as, x bar transpose A transpose A x bar plus, we know that y transpose y bar transpose A x bar and x bar transpose A transpose y bar these are equal to each other, because these are scalar quantities one is a transpose of the other. So, I can simply going to write this is as minus 2 x bar transpose A transpose y bar plus y bar transpose y bar plus the regularization factor lambda x bar transpose x bar and this is your F, F of x bar.

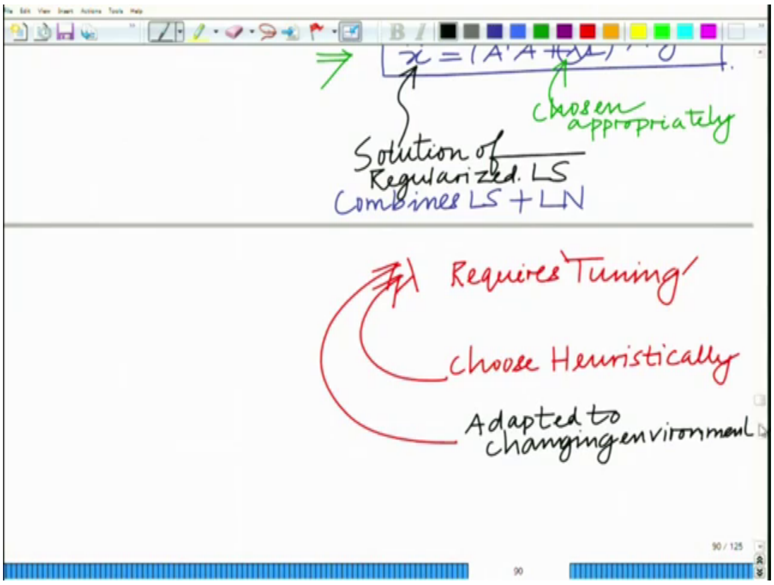
(Refer Slide Time: 11:36)



Now if you take the gradient of this with respect to x bar, what that gives us is well twice a transpose is a symmetric. So, this is twice x bar transpose p x bar the derivatives twice p x bar. So, this is twice A transpose x bar minus twice x bar transpose c bar derivative is c bar derivative with respective or gradient with respective. (Refer Time: 12:06) So, this will be minus twice A transpose y bar plus derivative of y bar, this does not depend on x.

So, its derivative with respect x bar is 0 plus, now this you can write as lambda x bar transpose identity matrix into x bar. So, this will be derivative will be lambda twice identity matrix into x bar. And now we said this equal to 0 to find the optimum. So this implies, now the factors of 2 will cancel and further you can write this as A bar A transpose A plus identity matrix or in fact, lambda times identity matrix into x bar equals A transpose y bar and this implies, that if you look at this solution, this will be A transpose A or this will be x hat you can call this as x hat equals A transpose A plus lambda times, I inverse ok

(Refer Slide Time: 13:30)

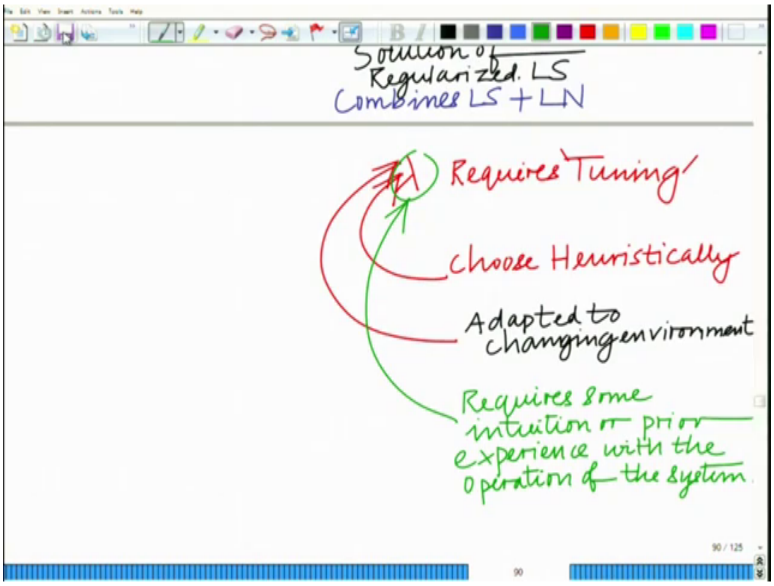


And therefore, this is basically the solution to the, your regularized this was that I choose a trade of between. So, it solution to your regularized L S, regularized S bar and combines both the criteria combines (Refer Time: 13:50) are least square as you can see the combines the properties of both the least squares and the least node ok, L S plus L N least square plus least node, it is the solution of the weighted optimisation problem.

Now a brief note regarding this lambda, this is the regularization parameter, now this needs to be chosen appropriately. Now remember the solution is easy, but we are not spoken of or we are not talked about, how to choose this? So, this lambda which is the regularising parameter this has to be chosen appropriately or this lambda requires tuning, what is known as tuning. In sense there one has to play around with this a little bit, to get the best solution alright.

So, you have to sort of adaptively, you might need to adaptively change this to get the best to at the, obtain the best solution. So, this can be chosen either you choose heuristically or this can be adaptively chosen adapted, adapted to changing environment and typically one needs to have a slight intuition about the system, that is this based on prior experience that is a this lambda can be determined over time, based on observing the system or based on observing the model for several time instances and then eventually determine, what is the lambda? What is the value of lambda? That best suits or what is the value of the regularization parameter that gives the best solution alright. So, that is the tuning process ok.

(Refer Slide Time: 16:12)



So, this requires basically tuning or either choose heuristically or tune it a appropriately, alright by adapting it to the changing by adapting it to a changing environment and it requires some intuition into the requires, some intuition or prior knowledge, we could put it that way or prior experience with the operation of the system.

So, as we determined the lambda, that is the regularizing parameter, that you will find the best solution alright. So, that basically completes our discussion on the regularised least squares, alright which combines both the least squares and the least norm flavours of this flame.

Thank you very much.