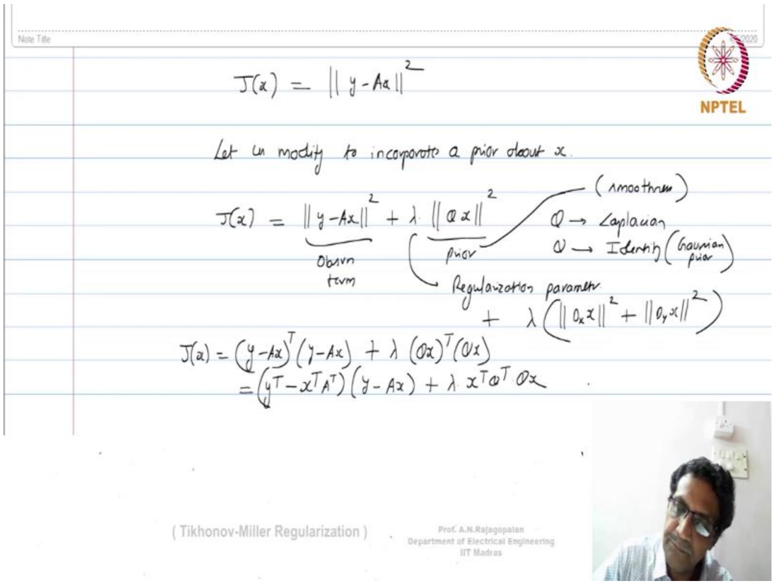
Image Signal Processing
Professor A. N. Rajagopalan
Department of Electrical Engineering
Indian Institute of Technology, Madras
Lecture 77
Tikhonov-Miller Regularization

(Refer Slide Time: 00:29)



Note Title

$$J(x) = \|y - Ax\|^2$$

Let us modify to incorporate a prior about x.

$$J(x) = \underbrace{\|y - Ax\|^2}_{\text{Obsvn term}} + \lambda \underbrace{\|Qx\|^2}_{\text{prior}}$$

(smoothness)

$Q \rightarrow$ Laplacian

$Q \rightarrow$ Identity (Gaussian prior)

Regularization parameter

$$+ \lambda \left(\|Q_x x\|^2 + \|Q_y x\|^2\right)$$

$$J(x) = (y - Ax)^T (y - Ax) + \lambda (Qx)^T (Qx)$$
$$= (y^T - x^T A^T)(y - Ax) + \lambda x^T Q^T Qx$$

(Tikhonov-Miller Regularization)

Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras

So, the idea is to really understand what do we really mean by Regularization, regularizing a solution. Now, if we revisit the least squares problem where we had J of x and we had a norm of what was that we had a y minus Ax square what we solved and then (())(0:39) square solution you got some answer.

Now let us modify it to incorporate a prior, incorporate a prior about x, now such a prior is normally very, very well known in the sense that these are generic priors, these are not very image specific in the sense that you are not trying to do it specifically for a certain class of image or something texture or phase or something this is a generic prior, what do we mean by that? We say that we modify J of x such that it becomes norm y minus Ax square plus lambda times norm Q x square. Suppose, this is called a prior and this is the other observation term, this is your observation term and the lambda is called a regularization parameter.

Now, what is this all mean? Prior to understand what we mean by how this actually brings instability, how this helps improve the condition number and so on, prior to that let us first understand what is this Q acting on x. Now, this Q could be an identity, now Q typically will be some kind of smoothness, this will signify the operator itself will not be a smooth operator but then this whole prior, let me call not Q but rather let me say that the prior is typically a smoothness prior, smoothness what does what mean?

That means locally intensities do not change rapidly, locally intensities ought to look somewhat similar so what is called a Markovian prior we have seen earlier also this is a very generic thing like any image any kind of natural image that see around you will find that the intensity is within a local region with some what looks similar, they are actually expected to look similar.

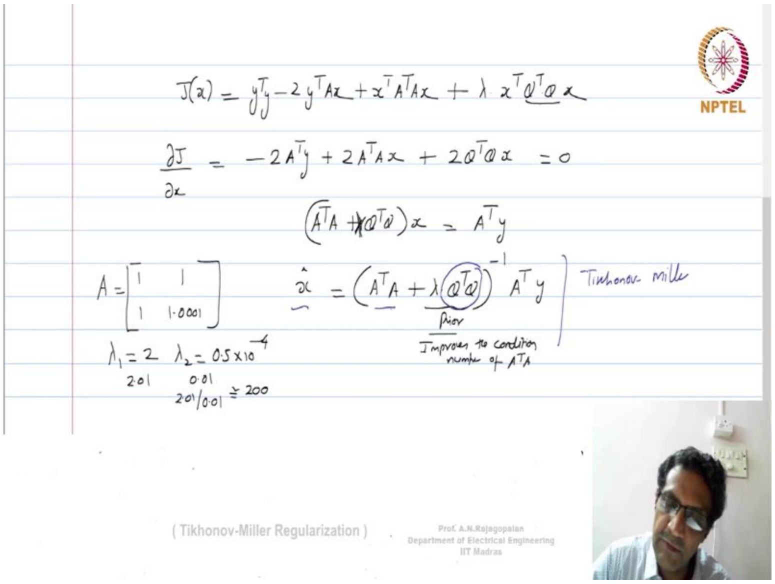
So, this Q will actually enforce that so one form of Q could be that it could be a Laplacian operator which then means that it (())(2:46) the x that you are trying to estimate should not only fall in the observational model but will follow the prior to some extent depending upon lambda which is the weighting parameter because Q times x will give you a Laplacian of the image which is like how the second gradient is changing and because we are trying to minimize the effects by saying that the energy in the Laplacian should be low, that means you are trying to say that the image should be low and smooth.

You can also go for Q to be identity in which case all that you will have is norm x square that will be like a simple Gaussian prior, that is also possible you can go for the Gaussian prior, you can also go for other forms of Q, you can have instead of one term you can have actually two terms here and one could be like norm of first derivative with respect to x of okay square and then plus dy second derivative with respect to square and then you can multiply this by lambda and then you have the observations term here, you can have various forms for this.

The whole idea is that you are trying to bring in a prior, prior knowledge about x because you believe that even without seeing the observation model I know for a fact that my x is locally smooth because what you are incorporating, if you know more about x you can always throw all that in into this framework.

Now, let us go back and do what we did before let us try to solve dou J by dou x, let us do dou J by dou x. Now, J itself and let us first write it down let us first expand J, if you expand here what do you get for J, so J of x is equal to y minus Ax the whole transpose into y minus Ax plus lambda Qx transpose into Qx or in other words is equal to y transpose minus x transpose A transpose into y minus Ax plus lambda x transpose Q transpose Qx.

(Refer Slide Time: 04:56)



$$J(x) = y^Ty - 2y^TAx + x^TA^TAx + \lambda\, x^T\underline{Q^TQ}\,x$$

$$\frac{\partial J}{\partial x} = -2A^Ty + 2A^TAx + 2Q^TQx = 0$$

$$(A^TA + \lambda Q^TQ)x = A^Ty$$

$$\underset{\sim}{\hat{x}} = (\underline{A^TA} + \underbrace{\lambda \boxed{Q^TQ}}_{\text{Prior}})^{-1} A^Ty$$

Tikhonov-Miller

Improves the condition number of A^TA

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$$

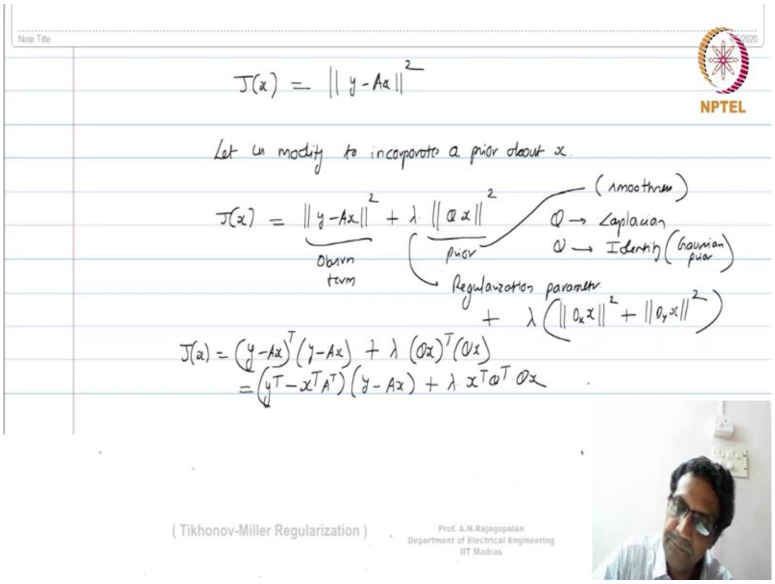
$\lambda_1 = 2 \quad \lambda_2 = 0.5 \times 10^{-4}$

$2.01 \quad 0.01$

$2.01/0.01 \doteq 200$

(Tikhonov-Miller Regularization)

Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras



Note Title

$$J(x) = \|y - Ax\|^2$$

Let us modify to incorporate a prior about x.

$$J(x) = \underbrace{\|y - Ax\|^2}_{\text{Obsvn term}} + \lambda \underbrace{\|Qx\|^2}_{\text{prior}}$$

(smoothness)

$Q \rightarrow$ Laplacian

$Q \rightarrow$ Identity (Gaussian prior)

Regularization parameter

$$+ \lambda\left(\|Q_x x\|^2 + \|Q_y x\|^2\right)$$

$$J(x) = (y - Ax)^T(y - Ax) + \lambda (Qx)^T(Qx)$$
$$= (y^T - x^TA^T)(y - Ax) + \lambda\, x^TQ^TQx$$

(Tikhonov-Miller Regularization)

Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras

Or in other words J of x is equal to from here and what do you find? You find it is y transpose y, first term is y transpose y, the second of course, all these are scalars, J of x is just a number minus 2y transpose Ax as we did before minus 2 y transpose Ax minus 2y transpose Ax plus x transpose A transpose Ax, this is what we had even earlier x transpose A transpose Ax so when they did least squares if you remember but now we have an additional term coming in the form of a prior which is lambda x transpose Q transpose Qx you see here that is what it is.

Now, taking though J by dou x will give you minus 2 A transpose y that is 2 A transpose Ax now you guys are familiar with respect to what to do and this is again a symmetric matrix, therefore it will be Q, 2Q transpose Qx is equal to 0 or in other words your A transpose A plus Q transpose Q the whole into x, x has multiply from the right it is a vector is equal to A transpose y or in other words x hat is equal to A transpose A plus Q transpose Q there should be gamma here, lambda here I forgot to put that lambda.

So, that should be lambda the regularization parameter should be there this Q transpose Q the whole inverse A transpose y, so now if you see the solution to compare the solution with the least square solution with you had let the A transpose A inverse A transpose y now you have an additional term coming here.

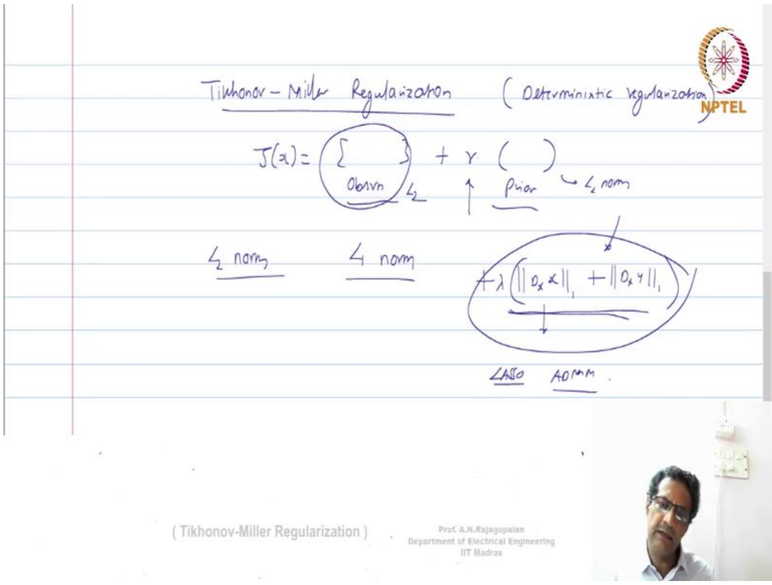
Now this prior, this is your prior the Q is a prior Laplacian or whatever, now this prior improves the condition number of A transpose A, improves the condition number of A transpose A, if you go back to the example that I gave you which was a very, very bad and kind of an example where A was very ill condition you had 1, 1, 1, 1.0001 you could have said that now I can add some increase to the Eigen value because we know the Eigen value spread was very high.

So, you could have said why not like this arbitrary add some values and sort of decrease for example, your lambda 1 is equal to 2 and then lambda 2 you had was 0.5 into 10 power minus 4, now we could have said something like why not we add you know let us say 0.01 to both so that would have meant this would have become 2.01, this would have become 0.01 approximately and this and the kappa would have been 2.01 by 0.01 it should be roughly 200, so we could have brought down the spread drastically but it is adding some number.

But the point is how do we add this numbers? When we cannot arbitrarily add something and we cannot simply make as my matrix table but simply arbitrary adding some entries but here if you see 2 A transpose A we are adding something but what we are adding is not all is not arbitrarily, we are bringing in a prior in the form of this matrix Q transpose Q multiplied by lambda this is the regularization parameter and now something sensible is being added to A transpose A in order to make it stable.

So, even usually A transpose A was not invertible now even it is even letting invertibility to A transpose A, it is improving its condition number and so on in order to be able to solve for x hat that is now a regularized solution and this kind of a regularization is called Tikhonov-Miller regularization. In fact let me write it down at all in a kind of a better manner.

(Refer Slide Time: 08:50)



Tikhonov – Miller Regularization (Deterministic regularization)

$J(x) = \underbrace{\{\quad\}}_{\text{Obsrvn}}{}_{L_2} + \gamma \, (\quad) \underset{\text{Prior}}{} \hookrightarrow L_2 \text{ norm}$

L_2 norm L_1 norm

$+\lambda \left(\|D_x x\|_1 + \|D_y y\|_1 \right)$

LASSO ADMM.

(Tikhonov-Miller Regularization)

Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras

Now you have in general what is called Tikhonov-Miller regularization, the idea of bringing in a trade-off between so you have a cost that is of the form of an observation term plus some gamma times a prior where you can control it, so it is like saying so regularization I am also saying that if I have a lot of confidence in my observation the sense of my noise is very little then I would go with a low gamma because then I do not have to use my prior too much.

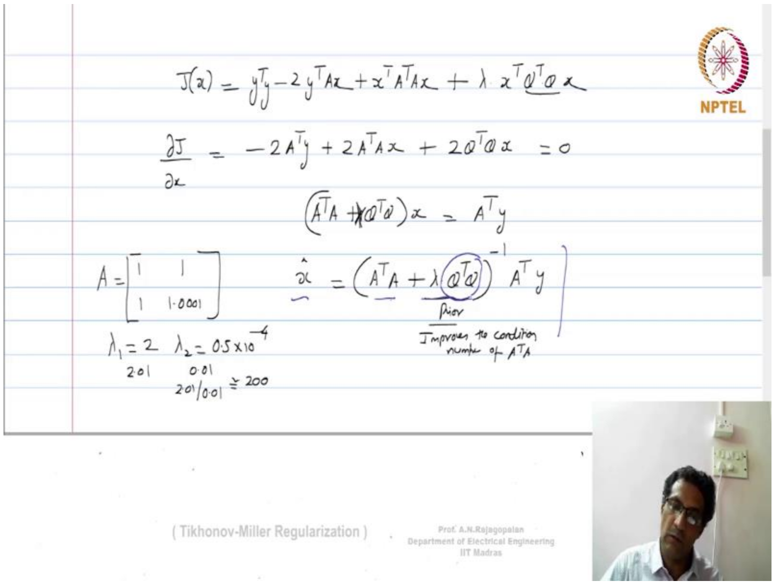
But on the other hand if I have a very noisy observation I might want to increase my gamma because then I will bank more on prior and so this kind of a trade-off between how much is the observation do you want to use and how much of the prior they want to use what other little weightage you want to give to both of them in order to be able to write at the x hat that is reasonably acceptable, it is what is called a regularization theory and this is called Tikhonov-Miller regularization and this comes in direct deterministic.

So, if you see that what we have done is really a deterministic regularization, there is a way to do stochastic regularization also in fact it is interesting that these two areas actually meet somewhere under certain conditions but stochastic is more general then a deterministic regularization and even this kind of regularization you can have in fact like different kinds of terms here for example, we have restricted ourselves L2 norm but then it is also very common to use an L1 norm and L1 norm is also possible to use for example, you can have plus when I said the gradient of x you could have used in used in L1 norm lambda times, the norm Dx y and then you could have used L1 norm what is also called a variation prior and so on.

The idea behind using an L1 norm for the prior was that really expect the natural gradients, images of natural, the gradients of natural images at quickly sparse so which has been observed and therefore it using a sparsity on the gradient of the image with some prior that let say people know they plotted several natural gradients they have plotted and they found that the gradients of naturalness is that typically sparse and therefore it is we bring in something like this then this can help actually preserve your edges even higher in a kind of a superior manner as compared to having a prior which is a L2 norm something like the ones that we saw on Qx square and so on.

But then the ability of L2 norm to keep the edges intact is not as good as the L1 norm but there are optimization algorithms that one can use in order to solve an L2, L1 combination, lasso is one such thing it uses what is called a ADMM, ADMM method we will not get actually go to these features of these methods but the point is once you fail we often fail the cost function then there are these optimization methods available out there which can directly follow it in order to be able to solve these problems, this is called Tikhonov-Miller regularization.

(Refer Slide Time: 12:08)



$$J(x) = y^Ty - 2y^TAx + x^TA^TAx + \lambda\, x^T\underline{Q^TQ}x$$

$$\frac{\partial J}{\partial x} = -2A^Ty + 2A^TAx + 2Q^TQx = 0$$

$$(A^TA + \lambda Q^TQ)x = A^Ty$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$$

$$\underset{\sim}{\hat{x}} = (\underline{A^TA} + \lambda \underbrace{Q^TQ}_{\text{Prior}})^{-1} A^Ty$$

Improves the condition number of A^TA

$\lambda_1 = 2 \quad \lambda_2 = 0.5 \times 10^{-4}$

$2.01 \quad 0.01$

$2.01/0.01 \simeq 200$

(Tikhonov-Miller Regularization)

Prof. A.N.Rajagopalan
Department of Electrical Engineering
IIT Madras

And I just wanted to wanted to say that you should be able to appreciate the fact that the prior information that enters in a certain form in order to be able to improve the stability of your final solution, that is in fact the goal after showing of this kind of regularization theory, we will as a follow-up we will see what is stochastic regularization and try to see what is the relation it has with respect to a deterministic regularization.