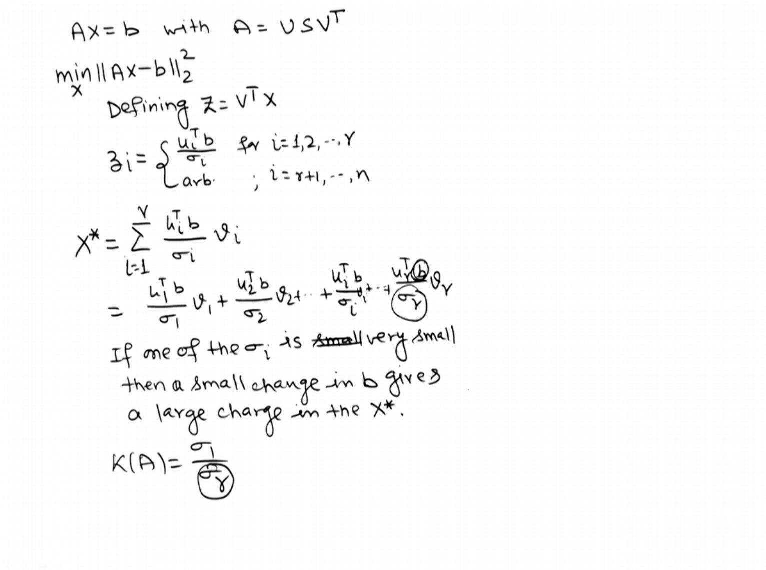
Matrix Analysis with Applications
Dr. Sanjeev Kumar
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 31
Regularization of Ill-Conditioned Systems

Hello friends. So, welcome to the lecture on Regularization of Ill-Conditioned Systems. So, in the last lecture we have learnt about the ill-conditioned systems and we have seen if the conditional number of the coefficient matrix is quite large then system is ill-conditioned. And if we have solve such system as such then the solution is not that useful because we are having a small error that may be due to sensor or due to computer we can get a very large deviation in our final solution. So, we are not sure about the solution whether the solution is correct or not.

So, in this lecture we will learn regularization of ill-conditioned systems means how to solve ill-conditioned systems by defining a suitable regularization term.

(Refer Slide Time: 01:31)



So, if you recall last to last lecture there we have learnt if we are having a system AX equals to b with singular value decomposition of A as U into S into V transpose then least square solution of this system is defined as AX minus b, then by making some calculation and defining Z equals to V T X. We have seen that ZI equals to U I T b upon sigma I for I equals to 1 2 up to r and arbitrary for I equals to r plus 1 up to n.

Later on by seeing the that the norm of Z equals to norm of X we have written the least square approximation of X as summation I equals to 1 to r where r is the number of nonzero singular values U I transpose b upon sigma I into V I. Now, just look here it will become u one transpose b upon sigma one into V 1 U 2 transpose b upon sigma 2 into V 2 plus sum U I transpose b upon sigma I plus up to U r transpose b upon sigma r, and here V I and here V r.

Now, if any 1 of sigma I if 1 of the sigma I is small or I will say is very small then a small change in b gives a large change in the solution X star this is why because b is numerator and sigma I which is very small close to 0 is in the denominator. And if you will divide something with a very small value which is very close to 0 you will get a large change and that is why that conditional number is sigma one upon sigma r means the biggest singular value upon a smallest singular value, and if it is very small this is quite large and system is ill-condition. And the same kind of analysis we can make here if sigma r is very small if you make a small change in b you will get a large change in X star means your system will be ill-conditioned.

Now how to avoid this problem of ill-conditioned system, that we will learn in this particular lecture.

(Refer Slide Time: 05:18)

Regularization of ill-conditioned system

Consider an ill-conditioned system $Ax=b$, where $\frac{\sigma_1}{\sigma_r} >> 1$, Then it might be useful to consider the regularized least square solution instead of the earlier one.

$$\min_{x} \frac{1}{2}\|Ax-b\|_2^2 + \underbrace{\frac{\lambda}{2}\|x\|_2^2}_{\text{regularization term}}$$

Thikonov regularization

Here, $\lambda > 0$ is called the regularization parameter which is not known apriori and has to be determined based on the problem data.

So, regularization of ill-conditioned system so, consider an ill-conditioned system AX equals to b where sigma one over sigma r is quite large where sigma 1 sigma 2 sigma r is the smallest singular value of a then it might be useful to consider the regularized least square solution instead of the earlier 1 here earlier 1 means which we have learnt in previous lectures. So, this solution is defined as you just find out X which minimize 1 by 2 times norm of X minus b plus here we are having a regularization term.

So now, instead of minimizing this only which we have taken in the earlier case we are minimizing this whole thing what will happen when you will minimize the norm of X it will it will stop to have a large deviation in the X and hence your solution will not be having the large change. So, this particular term is called regularization term and this kind of regularization is called Thikonov regularization which is quite popular in regularization theory. So, here I am telling it in terms of linear systems in terms of matrix analysis.

So, here this particular parameter lambda will be a positive real number. And it is called the regularization parameter which is not known a priori and has to be determined based on the problem data we will learn how to determine it by taking an example.

(Refer Slide Time: 09:39)

Now,

$$\min_X \frac{1}{2}\|AX-b\|_2^2+\frac{\lambda}{2}\|X\|_2^2 = \min_X \left\|\begin{pmatrix} A \\ \sqrt{\lambda} I\end{pmatrix}X-\begin{pmatrix} b \\ 0\end{pmatrix}\right\|_2^2 \quad — \quad (2)$$

For $\lambda>0$, the matrix $\begin{pmatrix} A \\ \sqrt{\lambda} I\end{pmatrix} \in \mathbb{R}^{(m+n)\times n}$ has always full rank n. the regularized system (2) has a unique solution.

Now, the normal equation to regularized system (2) can be written as

$$\begin{pmatrix} A \\ \sqrt{\lambda} I\end{pmatrix}^T\begin{pmatrix} A \\ \sqrt{\lambda} I\end{pmatrix}X=\begin{pmatrix} A \\ \sqrt{\lambda} I\end{pmatrix}^T\begin{pmatrix} b \\ 0\end{pmatrix}$$

or $(A^TA+\lambda I)X = A^Tb$

$\Rightarrow$ If the SVD of $A = USV^T$, then

$$[(USV^T)^T(USV^T)+\lambda VV^T]X=(USV^T)^Tb$$

$$(V(S^TS+\lambda I)V^T)X = VS^TU^Tb$$

$$(S^TS+\lambda I)V^TX = S^TU^Tb$$

Put $V^TX=Z$

$AX=b$

$A^TAX=A^Tb$

$X=(A^TA)^{-1}A^Tb$

$(USV^T)^T(USV^T)$

$(VS^T\underline{U^T)(U}SV^T)$

VS^TSV^T

Now, we have to minimize this particular function let us say this equation 1. So now, minimization of half norm AX minus b plus lambda by 2 the norm of X and will be equals to I can write this minimization of a root lambda I into X minus b 0. So, both of these are similar. So, let us say it equation 2.

Now for lambda greater than 0, so if lambda is non 0 what will happen the size of matrix A root lambda I will be means it will be having m plus n number of rows m from a n from I, because I is a n by n identity matrix into n number of columns and has always full rank full rank means n because even though if a is 0 I is a n by n matrix. So, the rank of this matrix will be always n and you are having a rank and coefficient matrix and x is having n number of unknowns. So, you will be always having a unique solution.

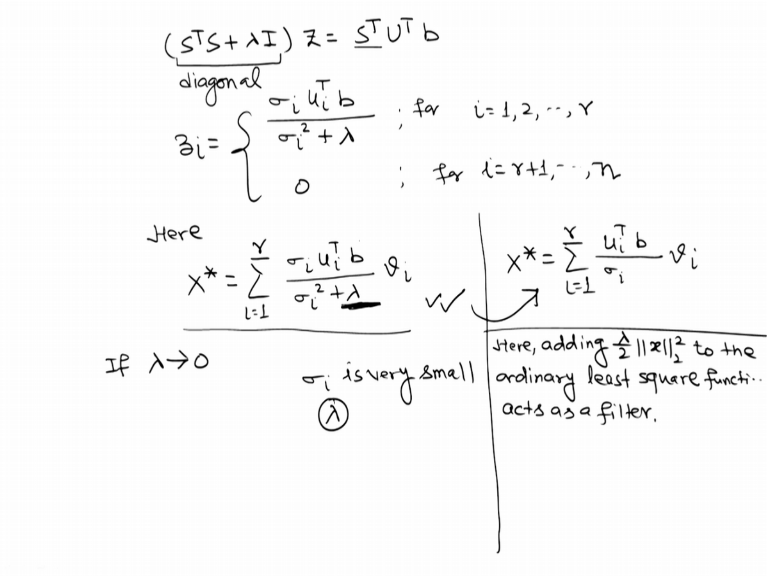
So, the regularized system 2 has a unique solution always. So, now how to find out this unique solution? So, we have learnt in a lecture if you are having a system AX equals to b where A is a rectangular matrix then the least square approximation of this A will be A transpose A into X equals to a transpose b, and from here X will be written as a transpose A inverse into A transpose into b which is the pseudo inverse of A.

So, in the same way now the normal equation so, this is called a normal equation to this system so, the now the normal equation to regularized system 2 can be written as a root lambda I transpose because like AX into equals to b. So, here A is replaced by A root lambda I into a root lambda I into X equals to A root lambda I transpose into b 0 or this can be written as a transpose into A plus lambda I X.

So, this is the left hand side equals to A transpose into b ok. If I take the singular value if the SVD of A equals to U S V transpose then this I can write U S V transpose transpose into A. So, A will become U S V transpose plus lambda I, I can write V into V transpose time X equals to U S V transpose into b. So, what I have done where ever I am having A in this normal equation I am writing it is singular value decomposition.

So, from here if I write this U S V transpose transpose into U S V T. So, this can be writ10 as v s transpose into U transpose because ABC transpose is C transpose b transpose a transpose into U S V T. So, U 2 into U will become I because U is an orthogonal matrix. So, this can be written as V S T S into V T. So, this system I can write V S T S plus lambda I into V transpose and this into X equals to what I am having here V S T U T b. Now if I premultiply this particular equation by a matrix V transpose then this will become V transpose into V equals to I. So, S T S plus lambda I into V transpose x equals to V transpose into V will become I S t U T b now put V T X equals to Z as we have done the earlier case.

(Refer Slide Time: 16:20)



So, then it can be written as S T S plus lambda I into Z equals to S T U T b, now if you look at this matrix it is a diagonal matrix because S S is a matrix which is having non 0 values only at few of the diagonal. So, S T S will be A square matrix having non 0 values if they are at the diagonal main diagonal only. So, then the solution Z I can be given as here sigma I due to this s transpose then U I transpose into b upon S T S will become sigma I square plus lambda for I equals to 1 2 up to r and it will be 0 for I equals to r plus 1 up to n.

Here, if I calculate the least square approximation in the same way which I have done earlier it will become summation I equals to one to r if I is having r non singular values sigma I U I transpose b upon sigma I square plus lambda into V I now in the earlier case where we have not use the regularization term I obtain this solution as I equals to 1 to r u I T b upon sigma I into V I and in the regularize case I am obtaining this solution. Now if lambda is tending to 0. In this case this particular solution put lambda equals to 0. So, sigma I will cancel sigma I square this will become this particular solution.

Means lambda equals to 0 means we do not have the regularization term. So, this is the same solution, but with the regularization term now how this solution is useful even though if a sigma I is very small ok, you can find out a suitable lambda so, that if you make a small change in b it will protect to having a large change in the final solution due to the choice of lambda. So, here what I can say. So, what I can write here adding lambda by 2 means this regularization term to the ordinary least square. So, this term acts as a filter means contribution from singular values which are larger relative to the regularization parameter means the sigma those are greater than lambda are left almost unchanged means you will get this kind of solution.

And where as you are having small sigma when compare to the lambda the solution will be this particular complete term will be treated as 0.

(Refer Slide Time: 20:47)

$$\frac{\sigma_i (u_i^T b)}{\sigma_i^2+\lambda} \approx \begin{cases} 0 & \text{if } 0 \approx \sigma_i << \lambda \\ \frac{u_i^T b}{\sigma_i} & \text{if } \sigma_i >> \lambda \end{cases}$$

How to choose λ,

Suppose the data are $b = b_{ex} + \Delta b$

$$X_{ex} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$ [Minimum norm solution of ordinary LLS with unperturbed bex]

But we can only compute $b = b_{ex} + \Delta b$, Since we don't know b_{ex}

Now the solution of regularized LLS is

$$X^* = \sum_{i=1}^{r} \left(\frac{\sigma_i (u_i^T b_{ex})}{\sigma_i^2+\lambda} + \frac{\sigma_i (u_i^T \Delta b)}{\sigma_i^2+\lambda} \right) v_i$$

So, I can write this particular term as sigma I U I t b upon sigma I square plus lambda this can be treated like as a filter and it will be 0 if you are having a singular value which is very close to 0 small singular values. So, there is no contribution from that term in the solution and it will become U I T b upon sigma I means here ordinary least square if sigma I is greater than lambda hence we can protect our solution to have a large change, because the term which will give you large change we have made the coefficient of that term as 0.

So, there will be no contribution of that particular term now question arise how to choose lambda. So, here suppose the data are b equals to b ex plus delta b ok. So, b x is a data without any perturbation and delta b that the possible perturbation in my right hand side vector b. So, then I can write my X e x as I equals to 1 to r U I T into b E X upon sigma I into V I or this should be write as b. So, this is like these. So, this is the minimum norm solution of ordinary least square with unperturbed data right hand side vector. So, this will come in this form.

Now, but we can only compute b we do not know how much perturbation is there in the b that is b ex plus delta b since we do not know b ex. Now the solution of regularized linear least square problem is X star equals to I equals to 1 to r sigma I U I T b upon sigma I square plus lambda plus sigma I U I t into delta b upon sigma I square plus lambda into V I. So, here I have written regularize least linear least square b I have written as b ex plus delta b here.

(Refer Slide Time: 25:14)

→ Here, we observe that

$$\sum_{i=1}^{r} \frac{\sigma_i(u_i^T b_{ex})}{\sigma_i^2+\lambda} \to X_{ex} \quad \text{as } \lambda \to 0$$

→ On the other hand,

$$\frac{\sigma_i(u_i^T \Delta b)}{\sigma_i^2+\lambda} \approx \begin{cases} 0 & \text{if } 0 \approx \sigma_i << \lambda \\ \frac{u_i^T \Delta b}{\sigma_i} & \text{if } \sigma_i >> \lambda \end{cases}$$

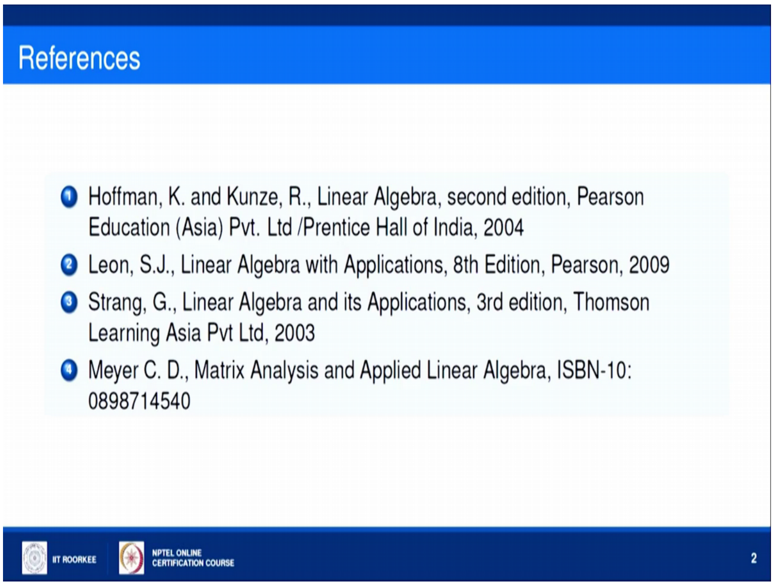
which suggests to choose λ sufficiently large to ensure that Δb in the data are not magnified by small singular values//

We observe that I equals to one to r sigma I u I t b ex upon sigma I square plus lambda tends to X ex as lambda tends to 0. Means whenever lambda is going to 0 your regularized linear least square solutions becomes the solution of the ordinary linear least square on the other hand we can see sigma I U I t into delta b means when perturbation in b upon sigma I square plus lambda will be approximately equals to 0 if sigma I is very very small to lambda and this is as the ordinary least square if sigma I means singular values are quite greater than lambda.

So, these 2 facts suggest us. So, we suggest to choose lambda sufficiently large to ensure that delta b in the data are not magnified by small singular values. So, in this way by choosing a suitable lambda if you know the singular values of the matrix A choose lambda somewhere which is greater than some of the small singular values, so that the contribution due to those small singular values will be neglected in the regularized least square approximation.

So, in this lecture we have learnt how to perform regularization or in this lecture we have learnt Thikonov regularization of the linear systems we have learnt how to choose regularization parameter. And we have done the analysis of regularize least square approximation based on the singular values.

(Refer Slide Time: 28:30)



These are the references.

Thank you very much.