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

Lecture - 35
Non Stationary Iterative Methods: Steepest Descent II

Hello friends. So, welcome to the second lecture on Non Stationary Iterative Methods. So, in this lecture we will continue the topic, which we have discuss in the previous lecture means steepest descent methods, we will see few more property of this gradient method. So, in previous lecture I told you that steepest descent method works like this.

(Refer Slide Time: 00:55)

$r_K = b - Ax_K$ (which is the search direction d_K)

$$\alpha_K = \frac{\langle r_K, r_K\rangle}{\langle r_K, Ar_K\rangle}$$

Theorem: In SD, consecutive search directions are orthogonal to each other ($d_K \perp d_{K+1}$) for $K = 0, 1, 2, \dots$

Proof: We have

$$\begin{aligned} d_{K+1} = r_{K+1} &= b - Ax_{K+1} \\ &= b - A(x_K + \alpha_K d_K) \\ &= \underline{b - Ax_K} - \alpha_K A d_K \\ &= r_K - \alpha_K A r_K \end{aligned}$$

Now

$$\begin{aligned} \langle d_{K+1}, d_K\rangle = \langle r_{K+1}, r_K\rangle &= \langle r_K, d_K\rangle - \alpha_K \langle Ad_K, d_K\rangle \\ &= \langle r_K, r_K\rangle - \frac{\langle r_K, r_K\rangle}{\langle r_K, Ar_K\rangle}\langle Ar_K, r_K\rangle \\ &= 0 \Rightarrow d_{K+1} \perp d_K \end{aligned}$$

You are having the residual as b minus AX K and which is the search direction also. Then what we have taken we have taken alpha K, that is the step length as rk transpose rk upon rk transpose A r k. Now, the first property we are going to look here which is a very important property in steepest descent method that is in steepest descent method consecutive search directions.

So, this is direction dk we denote it by dk in general setting are orthogonal to each other. Means, what I want to say that dk is orthogonal to dk plus 1 for k equals to 0 1 2 and so on, in d 0 is orthogonal to d 1 then d 1 is orthogonal to d 2 and so on. So, let us try to prove it. So, we have dk plus 1 which is basically rk plus 1 residual and this is b minus A X K plus 1 in gradient descent steepest descent.

This equals to b minus A and X K plus 1 will be X K plus alpha k dk. So, this becomes b minus A X K minus alpha is any scalar. So, alpha is a scalar. So, alpha k into A d K, b minus A X K can be written as rk minus alpha k A and dk is also rk. So now, if I check the inner product of dk plus 1 with dk which is inner product of rk plus 1 with rk this becomes rk dk minus alpha k rk; sorry A d k with dk, because dk plus 1 I am writing in this form.

So, d inner product of dk with dk plus 1 will be first term will come rk dk minus alpha k I am taking out A rk or I have written A dk here with dk. This becomes rk into rk so, I am writing everything in terms of rk, because dk equals to rk minus if you see the value of alpha k alpha k is inner product of rk with rk upon rk A rk, into A rk rk.

So, this will be 1. So, it will become 0. So, the inner product of 2 consecutive search directions are 0, means they are orthogonal to each other.

(Refer Slide Time: 05:48)

Nonsymmetric Steepest descent

In SD method, the matrix A must be symmetric and PD in order to have a unique minimum of functional

$$q(x) = \frac{1}{2} x^T A x - x^T b$$

which is also a solution of the linear system $Ax = b$.

Now, consider that A is not SPD, but it is non-singular. Then, the matrix $A^T A$ is SPD and the algorithm can be applied to the normal equation

$$A^T A x = A^T b$$

$$\underset{\substack{\downarrow \\ \text{SPD}}}{\hat{A}} x = \hat{b}$$

Now, let us see another variant of steepest descent method that is steepest descent method when the given matrix A is not symmetric as well as positive definite; so non-symmetric steepest descent. So, we have seen that in steepest descent method the matrix A must be symmetric and positive definite. In order to have a unique minima of functional q X equals to half X T A X minus X transpose b, which is also a solution of the linear system A X equals to b.

Now, just consider that A is not SPD. SPD stands for symmetric and positive definite, but it is non-singular. Then, the matrix A transpose A is symmetric as well as positive definite. And, the algorithm can be applied instead of AX equals to b we can apply it to the normal equation of AX equals to b and which is A transpose AX equals to A transpose b.

So, this I can write AX equals to b, where A cap is a transpose A and b cap is a transpose b. So, here you can easily see that A cap is symmetric and positive definite. So, I can apply the steepest descent method. So, this is the strategy for applying steepest descent method for a general system, where A is not symmetric and positive definite let us take an example of it.

(Refer Slide Time: 09:42)

Ex:. Solve the linear system $Ax=b$ using SD method with $x_0=(0,0,0)^T$ and

$$A=\begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 2 \\ 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad b=\begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}$$

Sol^n:

$$A^TAx = A^Tb$$

$$A^TA=\begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 2 \\ 1 & 1 & 0 \end{bmatrix}^T\begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 2 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 10 & 4 & 0 \\ 4 & 11 & 6 \\ 0 & 6 & 4 \end{bmatrix}; \quad A^Tb=\begin{bmatrix} 12 \\ 7 \\ 2 \end{bmatrix}$$

$$10>0, \ \begin{vmatrix} 10 & 4 \\ 4 & 11 \end{vmatrix} = 94>0,$$

$$\det(A^TA) = 16 > 0$$

Hence, A^TA is a SPD matrix.

$$r_0 = A^Tb - A^TAx_0 = (12,7,2)^T$$

$$\alpha_0 = \frac{r_0^Tr_0}{r_0^TAr_0}$$

Solve the linear system AX equals to b using steepest descent method with initial solution as 0 0 0 transpose and A is given as 3 1 0 0 3 2 1 1 0 and b is 4 1 0. So, if you see that the matrix A is neither symmetric nor positive definite. Hence, we cannot apply the gradient descent steepest descent algorithm directly on this system. Means, we cannot minimize the functional half of X T A X minus X transpose b using the steepest descent method. So, what we will do here we will apply the method on A transpose AX equals to A transpose b. So, let us first calculate A transpose A. So, A transpose A will become 3 1 0 0 3 2 1 1 0 transpose into 3 1 0 0 3 2 1 1 0 and this comes out to be 10 4 0 4 11 6 and 0 6 4.

The same time we calculate A transpose into b and it becomes 12 7 and 2. So, now, instead of the original system AX equals to b, we are going to solve A transpose AX equals to A transpose b. So, for applying the steepest descent method here this matrix would be positive definite, because it will be symmetric it is product of A matrix with it is transpose. So, it will be symmetric always.

So, positive definite if we check here so, 10 is greater than 0 if I take this 10 4 4 11. So, this is 94, which is greater than 0 and determinant of A transpose A comes out to be 16 which is again positive. So, hence A transpose A is A symmetric and positive definite matrix and we can apply the steepest descent method here.

So, let us apply here the method. So, my r 0 will become A transpose b minus A transpose A X 0, which comes out to be 12 7 2 transpose. Now, I calculate my alpha 0 which is r 0 transpose into r 0 upon r 0 transpose A r 0. So, it will come out to be A scalar and in the similar way as we have done in the previous lecture we can apply the steepest descent method.

(Refer Slide Time: 14:28)

This procedure is called the residual norm steepest descent.

Here, the functional being minimized is

$$\psi(X) = \frac{1}{2}\langle AX, AX\rangle - \langle X, A^T b\rangle$$

$$q(X) = \frac{1}{2}\langle X, AX\rangle - \langle X, b\rangle$$

This method minimizes the Euclidean norm of the residual $\|AX-b\|_2^2$.

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

And, the solution of this system will be the solution of the original system in particular this procedure; where the steepest descents we are applying on the normal equations of the original system. So, this procedure is called the residual norm steepest descent. Here, the functional being minimized is so, it will be instead of q X I am writing psi X, because it is A different function and it is half inner product of AX with ax minus X A transpose b.

If, you check the earlier one in case of extended steepest descent it was q X, which was half X with AX minus X b. So, here you can notice that we are having this A transpose b instead of b it is because, now my b is in the normal equation A transpose b. And, similarly instead of this X I am having ax here because it is I am applying of A transpose A instead of A.

And, this method minimizes the Euclidean norm of the residual that is AX minus b. And, if you can recall the least square approximation method in that method we have written the solution like X equals to A transpose A inverse into A transpose b. So, here this solution, which we are obtaining with residual norm steepest descent method is similar what we are obtaining using least square approximation.

Now, let us see another important property of the steepest descent method.

(Refer Slide Time: 17:40)

§ Instant convergence of SD method

Theorem: If the initial error is an eigenvector of A in the SD method, then the method converges in just ONE iteration.

Proof: Let (λ, v) be an eigenpair of A,

$$Av = \lambda v \quad \text{— (1)}$$

Now assume initial solution as $x_0 = x^* - v$, where x^* is the exact solution of $Ax = b$

So

$$r_0 = b - Ax_0$$
$$= b - A(x^* - v)$$
$$= b - Ax^* + Av = Av = \lambda v$$

$$\alpha_0 = \frac{\langle r_0, r_0 \rangle}{\langle r_0, Ar_0 \rangle} = \frac{\langle \lambda v, \lambda v \rangle}{\langle \lambda v, A\lambda v \rangle} = \frac{1}{\lambda}$$

$e_0 = r_0$
$= x^* - x_0$
$= x^* - (x^* - v)$
$= v$

So, this I will write as instant convergence of steepest descent method. So, let us write this result. So, if the initial error; that means, r 0 which is b minus AX 0 is an eigenvector of the coefficient matrix A in the steepest descent method, then the method converges in just one iteration only.

So, what I want to say if your initial solution you choose in such a way, that the initial residual or initial error becomes an eigenvector of the coefficient matrix. Then the steepest descent method will converge in just one iteration to the exact solution let us see the proof of this. So, let lambda V be an Eigenpair of A. So, lambda is an eigenvalue and corresponding eigenvector is V it means A into V equals to lambda into V let us say this is equation 1.

Now, assume initial solution as X 0, which is the error. So, this is X star minus V, where X star is the exact solution of AX equals to b. So, what I am assuming here I am taking the initial solution as the error, if I choose my initial solution in this way then what you can see from here that the error in the initial solution will be. So, initial error e 0 will be or r 0 here X star minus X 0, this is X star minus X 0 is X star minus V. So, this will become V.

So, I am taking initial error as the eigenvector of the matrix A. So, now, calculate r 0. So, calculate r 0 r 0 will become b minus AX 0. So, b minus A X 0 we have chosen X star minus V. So, b minus AX star plus AV AX star equals to b. So, b will be cancel out it will remain as AV and AV is nothing just lambda times V.

So, the initial residual is a multiple of scalar multiple of V, which is again an eigenvector same eigenvector here if we calculate alpha 0. So, alpha 0 is inner product of r 0 with r 0 upon inner product of r 0 with ar 0. So, this comes out to be lambda V into lambda V upon lambda V into A lambda V. So, lambda you can take out lambda AV and this comes out to be one upon lambda because AV again become the lambda V.

So, in the numerator you will be having lambda square in the denominator you will be having lambda q. So, it comes out to be 1 upon lambda. So, the step length is 1 upon lambda and initial residual is lambda times that eigenvector, now calculate X 1.

(Refer Slide Time: 23:29)

$$X_1 = X_0 + \alpha_0 r_0$$
$$= X_0 + \frac{1}{\lambda}(\lambda V)$$
$$= X_0 + V$$
$$= X^* \text{ (exact sol}^n \text{ of } AX = b)$$

Ex.:

$$A = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}; \quad b = \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} \qquad \underline{Ax = b}$$

Here $(\underset{\lambda}{2}, \underset{\vartheta}{(1, 1, 0)^T})$ is an eigenpair of A.

If we choose

$$X_0 = X^* - \vartheta = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}$$

$$r_0 = b - AX_0 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}; \quad \alpha_0 = \frac{1}{2}$$

$$X_1 = X_0 + \alpha_0 r_0 = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}$$

So, X 1 will become X 0 plus alpha 0 r 0, and what is my claim my claim is that the steepest descent method will converge in just 1 iteration. So, this X 1 should be equals to the exact solution which is X star in our case.

So, it will X 0 plus alpha 0 is 1 upon lambda r 0 is lambda into V. So, this comes out to be X 0 plus V and X 0 plus V is nothing just X star, because X 0 is X star minus V and which is exact solution of AX equals to b. So, in this way we have seen that if the initial error you choose the initial solution in such a way that the initial solution error is an eigenvector of the matrix A, then the steepest descent method will converge in just 1 iteration.

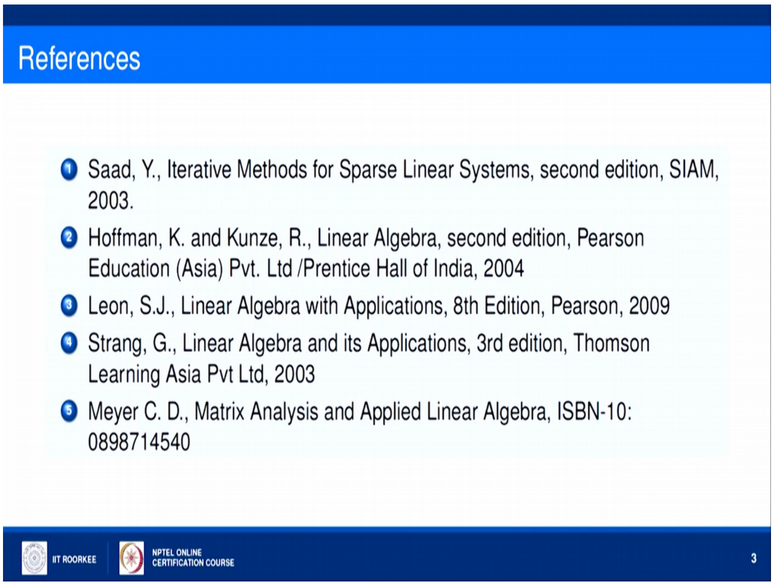
So, if somehow you are having an idea of the eigenvector of A you can choose your initial solution in such A way that the error in the zeroth iteration will be that eigenvector then your method will converge in just 1 iteration this is the idea. So, let us take an example based on this. So, consider A equals to 3 minus 1 1 minus 1 3 minus 1 and 1 minus 1 3. So, this is the same example which we have taken in the earlier lecture.

And, here b the same b I am taking minus 1 7 minus 7. So, solve AX equals to b using steepest descent. So, here if you see one of the eigenpair of A is 2 1 1 0. So, this is eigenvalue and this is corresponding eigenvector. So, is an eigenpair of A. If, we choose X 0 as X minus V, where X is the exact solution of this so, 1 2 minus 2 minus 1 1 0 so, it comes out to be 0 1 minus 2.

Then, my r 0 will become b minus AX 0 and b minus AX 0 means. So, b minus AX 0 this comes out to be 2 2 0. Here, if I calculate alpha 0 it will be 1 upon lambda so, 1 upon 2. So, what is X 1 X 1 is X 0 plus alpha 0 r 0 X 0 is 0 1 minus 2 plus 1 by 2 into 2 2 0, which is 1 2 minus 2 that is the same as the initials as the exact solution X star.

So, by this example we have verified the result, which is given in the theorem that you the steepest descent method will converge in just one iteration.

(Refer Slide Time: 27:38)



So, these are the references, for this lecture and in this lecture we have seen some properties of steepest descent. And then we have seen the residual norm steepest descent method for general system AX equals to b, where the matrix A is not A SPD matrix means symmetric and positive definite matrix.

In the next lecture we will learn another gradient method that is called conjugate gradient method, which is having faster convergence when compare to the steepest descent method.

Thank you very much.