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

Lecture - 36
Krylov Subspace Iterative Methods (Conjugate Gradient Method)

Hello friends. So, welcome to the 36th lecture of this course. So, as you remember in past couple of lectures I have discuss the steepest descent method which is coming from the class of non stationary method. Here we are going to look another type of non stationary method which is called Krylov subspace iterative methods.

So, before going to the methods I will introduce you what we mean by Krylov subspace. So, let me write the definition of Krylov subspace.

(Refer Slide Time: 01:0 6)

Def^n. Krylov subspace:

Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. Then, the Krylov subspace $K_j(A,b)$ is defined as

$$\underline{K_j(A,b)} = \text{Span}\{\underline{b}, \underline{Ab}, \underline{A^2b}, \cdots, \underline{A^{j-1}b}\} \subseteq \mathbb{R}^n$$

Ex.: Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

then $Ab = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = A^2b = A^3b \cdots$

$K_1(A,b) = L[(1,-1)^T]$

$K_2(A,b) = L\{b, Ab\} = L[(1,-1)^T]$

So, let A be a n by n matrix having real entries and b be a column vector having dimension n. Then the Krylov subspace which is denoted as K j A b is defined as, it means the j-th Krylov subspace of A and b is the linear span of the vectors b, A into b, A square into b and in this way up to a raised to the power j minus 1 into b. And you can observe that again this Krylov subspace is a subspace of R n because all these vectors are having n components. So, all these vectors are coming from the vector space R n. Let us take an example of it. Let A equals to 2 1, 1 2 and b equals to 1 and minus 1 then if I calculate A b it will be 2 1, 1 2 multiplied with column vector 1 minus 1 and it comes out to be 1 and minus 1 which is equals to A square b equals to A cube b and so on.

So, here Krylov subspaces that is if I write K 1 A b K 1 A b will be the linear span or let me denote this linear span by L of 1 minus 1 transpose. Similarly K 2 A b will be linear span of vector b and Ab. So, this will be linear span of b and A b are the same vector here; so 1 and minus 1 transpose. So, in this way we can define the Krylov subspace.

(Refer Slide Time: 04:52)

Krylov matrix: Let $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^{n}$, then Krylov matrix of j^{th} order of A and b is

$$\begin{bmatrix} b & Ab & \cdots & A^{j-1}b \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

Next I will define Krylov matrix. Again let A be n by n matrix having real entries and b is a column vector having n component. Then Krylov matrix of j-th order of A and b is the matrix having first column as b, the second column will become A b and in this way the j-th column will become A raised to power j minus 1 into b. So, in this way if I am having if I want to define nth order, so this j equals to n and in that case it will become a square matrix of order n.

Next what is the motivation of using Krylov subspace for solving the iterative system?

(Refer Slide Time: 0 6:1 6)

Motivation. Consider the matrix

$$A = \begin{bmatrix} 0 & 2 & 1 \\ -1 & 3 & 1 \\ -2 & 2 & 3 \end{bmatrix};$$ the characteristic polynomial of A is $c_A(\lambda) = \det(A - \lambda I)$
$$= \lambda^3 - 6\lambda^2 + 11\lambda - 6$$

Now, by the Calley Hamilton theorem, we have

$$A^3 - 6A^2 + 11A - 6I = O_{3\times 3} \checkmark$$

or $$A^{-1} = \frac{1}{6}(A^2 - 6A + 11I) = p(A) \checkmark$$

Now, if we have a system $AX = b$, where A is an invertible matrix, then

$$\begin{aligned} X &= A^{-1} b \\ &= \frac{1}{6}(A^2 - 6A + 11I)\, b = p(A)\, b \\ &= \frac{1}{6}[A^2 b - 6Ab + 11b] \\ &= L(A^2 b, Ab, b) \\ &= \underline{K_3(A, b)} \end{aligned}$$

So, let us consider, so I want to talk about motivation. So, consider the matrix a which is a 3 by 3 matrix having entries 0 2 1, minus 1 3 1 and then minus 2 2 3. Now, the characteristic polynomial of a is C A lambda which is just determinant of a minus lambda I and if I calculate it comes out to be a polynomial in lambda of degree 3 which is lambda cube minus 6 lambda square plus 11 lambda minus 6.

Now, by the Calley Hamilton theorem we have. So, what Calley Hamilton theorem tells us? That every matrix satisfies its characteristic equation, so it means I can write a cube minus 6 A square plus 11 A minus 6 I, where I is an identity matrix of order 3 by 3 in this case equals to 0 which is a null matrix of size 3 by 3 or I can write from here if a is an invertible matrix then a inverse equals to 1 by 6 A square minus 6 A plus 11 I. So, what I have done? I have multiplied this characteristic equation means matrix in terms of matrix by A inverse and I have written in this way let us say this is a polynomial of matrix A.

Now, from here I can say that the inverse of A can be expressed in terms of as a polynomial of matrix A. Now, if we have a system AX equals to b, where A is an invertible matrix then the solution of this system can be written as X equals to A inverse into b which I can write in case of this matrix 1 by 6 A square minus 6 A plus 11 I into b which is nothing just p of A into b, if I do it what I will get 1 by 6 A square into b minus 6 A b plus 11 b.

If you see this particular vector that is a square b minus 6 A b plus 11 b it is a vector column vector having 3 components. So, 3 by 1 vector this is just linear span of A square b, A b and b which is means this vector can be written as in the linear combination of A square b, A b and b which is in this from and this is nothing just Krylov subspace which is K 3 A b. So, hence the solution of the system AX equals to b is a vector from the Krylov subspace K 3 A b. So, that is the motivation of using Krylov subspace for solving iterative system.

Now, what I want to mention here if A is a large and sparse matrix, but it is non singular then Krylov subspace method is a very effective method for solving linear system. We will get the solution with less computational complexity as compared to any other iterative method extended iterative methods like Jacobi and Gauss Seidel or some direct method. So, this is the motivation of using Krylov subspace method.

(Refer Slide Time: 12:26)

Main result: Let $AX = b$ be a linear system, where $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$. Then, there exists a real polynomial p

$$p(t) = \sum_{j=0}^{m-1} \alpha_j t^j$$

such that

$$X = p(A)\,b = \underline{\sum_{j=0}^{m-1} \alpha_j A^j b}$$

$$m = n$$

$$m(A) = 0$$

One more thing I want to write about, this that is a main result. So, we can summarize this main result as let AX equals to b be a linear system, where A belongs to R n by n and b is a vector from n tuple real vector space we do not care whatever large n. Then there exist a real polynomial p p given as p t equals to j equals to 0 to m minus 1 alpha j t j, where alpha 1, alpha 2 all these alpha 0 alpha 1 up to alpha m minus 1 all these are scalars, all these are real numbers basically such that the solution of system X equals to b can be written as X equals to p A into b and this can be written as j equals to 0 to m minus 1 alpha j A j b, now this is the p A.

Now, it is not necessary here that m equals to n, means the degree of this polynomial p t will be a characteristic polynomial of the matrix A. One of such case is the use of minimal polynomial because minimal polynomial of a matrix A is also 0 and from there we can use we can find out A inverse which is having pure degree sometimes compared to the characteristic polynomial and that is why I was saying that these methods are quite effective and having less computational complexity.

(Refer Slide Time: 15:12)

Revisiting Krylov subspaces

The Krylov subspace $K_j(A,b)$ is the column space of the Krylov matrix.

Now, we want to choose the best combination as our improved x_j. Now, there are various ways of defining this 'best'.

1. The residual $r_j = b - Ax_j$ is orthogonal to $K_j(A,b)$ (Conjugate Gradient method)
2. The residual r_j has minimum norm for x_j in $K_j(A,b)$ (GMRES method)
3. r_j is orthogonal to a different space $K_j(A^T, b)$ (Biconjugate method)
4. The error e_j has minimum norm (SYMMLQ)

Now, just try to formulate method using this sub spacing. So, let me revisit Krylov subspaces the Krylov subspace K j A b is the column space of the Krylov matrix which is quite obvious because the columns of Krylov matrix is nothing just the vectors b A into b A square into b up to A raised to the power j minus 1 into b.

Now, for solving a linear system we want to choose the best combination as our improved X j in j-th iteration if you are having X j minus 1. Now, there are various ways of defining this best or there are various ways of choosing X j in j-th Krylov subspace of A and b. So, let us see what are those way. So, the first way is the residual r j which is nothing just b minus A X j. So, choose X j in such a way from the Krylov subspaces K j A b such that the residual r j is orthogonal. So, r j should be an orthogonal vector to K j A b means, r j is orthogonal to all vectors from this Krylov subspace. If you use this strategy then we will have a method that is called conjugate gradient method or in short CG method.

The second strategy is the residual r j has minimum norm for X j in K j A b. So, if we follow this strategy then the method name is we got another Krylov subspace iterative method that is called GMRES method. So, this is gradient method for minimum with minimum residual. The other way of defining this space maybe r j which is the residual is orthogonal to a different space K j A T b. So, here you just notice that instead of a I am having Krylov subspace of A transpose and b this method is called biconjugate method. The next strategy is the error e j has minimum norm. So, the error in j-th iteration has minimum norm and this strategy we follow in a method called SYMMLQ. So, in this way we are having these 4 Krylov subspaces method based on defining this based in different way.

There are other methods also. In this course I will focus on this method because if you know one method you can follow the other methods from the literature in a simple manner.

(Refer Slide Time: 21:19)

Conjugate Gradient Method: It is one of the standard Krylov subspaces method which is used when the system is large and sparse, but the matrix A is symmetric and positive definite (SPD).

Algorithm

Input: A SPD matrix A, b, x_0

1- Compute $r_0 = b - Ax_0$; set $d_0 = r_0$

2- For $k = 0, 1, 2, \cdots$ until convergence; do

3- $\alpha_k = \dfrac{\langle r_k, r_k \rangle}{\langle Ad_k, d_k \rangle}$

4- Compute $x_{k+1} = x_k + \alpha_k d_k$

5- Compute $r_{k+1} = r_k - \alpha_k d_k$

6- If $r_{k+1} = 0$, then stop, else

7- $\beta_k = \dfrac{\langle r_{k+1}, r_{k+1} \rangle}{\langle r_k, r_k \rangle}$

8- Compute $d_{k+1} = r_{k+1} + \beta_k d_k$

9- End For.

So, let us learn; what is conjugate gradient method. So, it is one of the standard Krylov subspaces method which is used when the system is large and sparse, but the matrix that is the coefficient matrix A is symmetric and positive definite. Like in the case of steepest descent method, so it is as SPD in short.

So, let us go to the algorithm how this method executes. So, input is a symmetric and positive definite matrix A which is the coefficient matrix of the linear system AX equals to b, right hand side vector b and an initial guess it 0 because it is a non stationary iterative method. So, we need some initial solution.

In first step compute R 0 which is b minus AX 0 and it is same as in case of steepest descent. After computing this set the search direction in the initial iteration is d 0 equals to r 0 which is again similar to steepest descent. Now, for K equals to 0 1 2 until convergence do calculate alpha K that is the step length which is again same as in case of steepest descent. In initial iteration then it will change because later on we will make the change in the search direction then 4th step is compute X K plus 1 equals to X K plus alpha K d K now same iterative equation as we are having in steepest descent.

Next from here this method will differed from steepest descent method. Compute r K plus 1 which was b minus AX A into X K plus 1 in steepest descent, but here it will become r K minus alpha K d K. Then if r K plus 1 equals to 0 then stop else calculate beta K which is r K plus 1 r K plus 1 upon r K r K inner product of r K plus 1, r K plus 1 upon inner product of r K, r K. Then what you do compute d K plus 1 as means next search direction, r K plus 1 plus beta K d K and then end of your for loop. So, this is the complete algorithm for conjugate gradient method.

(Refer Slide Time: 27:15)

Explanation: The conjugate gradient algorithm returns approximations

$$x_j \in x_0 + \underline{K_j(A, r_0)} \quad \text{for } j = 0, 1, 2 \cdots$$

such that

$$\|x - x_j\|_A = \min_{q \in P_{j-1}} \|(I - Aq(A))(x - x_0)\|_A$$

where $\|x\|_A = \sqrt{x^T A x}$

First step: $x_0 \to x_1$ $\qquad \alpha_0 r_0 \in K_1\{A, r_0\}$

$$x_1 = x_0 + \underline{\alpha_0 r_0} \in x_0 + K_1$$

Here, $\alpha_0 = \dfrac{\langle r_0, r_0 \rangle}{\langle d_0^T A d_0} = \dfrac{\|r_0\|_2^2}{\|d_0\|_A^2} \neq 0$

$$r_1 = r_0 - \alpha A_0 d_0 \in K_2(A, b)$$

If $r_1 = 0$ STOP

Otherwise $\{r_0, r_1\}$ is an orthonormal basis of $\underline{K_2(A, b)}$.

If I need to explain this method theoretically then let us have a quick look on it. So, let us have a quick look on the explanation of this method in theoretical sense. So, the conjugate gradient algorithm returns approximations means in different iterations X j that is the solution in j-th iteration as a vector of the sum of initial solution plus a vector from j-th Krylov subspace of A and R 0. And this is true for j equals to 0 1 2 and so on, such that the error in j-th iteration with a norm A should be minimum over a vector q which is from a polynomial of degree j minus 1 I minus A q A into X minus X naught, where the norm A is defined as square root X T A X for a given vector A. So, it is the norm induced by the matrix A.

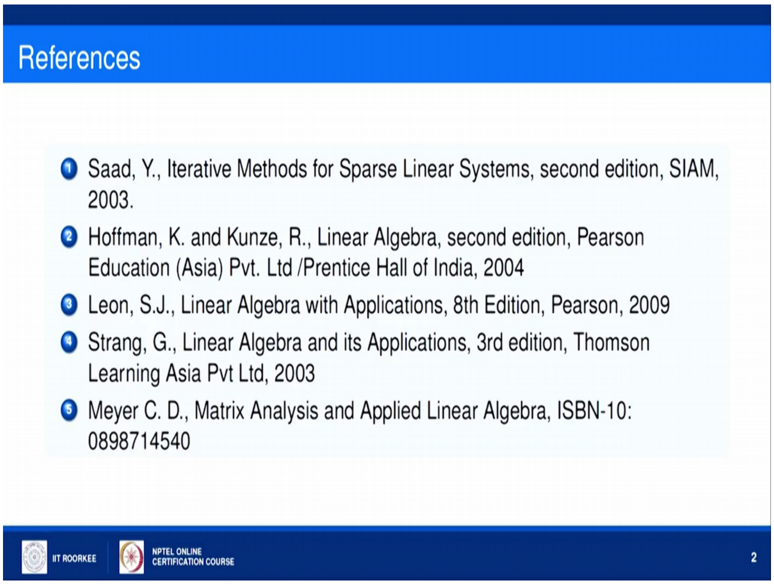
Now, if we look first step of this method means forgoing X 0 to X 1, what we are having X 1 equals to X 0 plus alpha 0 R 0. Now, this alpha 0 R 0 is a vector in K of 0 A of r 0 r 0 is alpha 0 r 0 is a vector in this Krylov subspaces because this Krylov subspaces will be having vectors r 0 means scalar multiple of r 0. So, it is X 0 plus this one.

Now, this belongs to X 0 plus K 1 sorry not 0. Here alpha 0 will become r 0 r 0 upon again if I write in this way d 0 T A d 0 and this is nothing just the square of the norm of r 0 upon a square of the norm induced by the matrix A on the vector d 0 and it cannot be 0 because r 0 is not 0. Once r 0 is 0 method will not proceed that is the termination or stopping condition for the conjugate gradient method.

Now, residual r 1 will be r 0 minus alpha A 0 d 0. Again this will become a vector in the second Krylov subspace of A and b. Here if r 1 is 0 then method will stop otherwise r 0 r 1 is an orthonormal basis of K 2 A b which is the condition when I told you about 4 methods that each time the residual r j will be the will be a vector which is orthogonal to the j-th Krylov subspace of A and r 0. So, in this way again when I will calculate r 2, then r 0, r 1 and r 2 will form an orthonormal basis for K 3 A b and so on. So, in this way we will proceed for subsequent iterations in conjugate gradient method.

So, in this lecture we have learnt the definition of Krylov subspace then we have seen the motivation of using Krylov subspaces for solving large and sparse linear systems. In the last phase of this lecture we have learnt the conjugate gradient algorithm. In the next lecture we will take an example of conjugate gradient method and then we will see some conditions for the convergence of this method. Later on in the next lecture we will learn preconditioning of conjugate gradient method.

(Refer Slide Time: 33:42)



These are the references for this lecture.

Thank you very much.