

# Foundation of Scientific Computing

**Prof. T. K. Sengupta**

**Department of Aerospace Engineering**

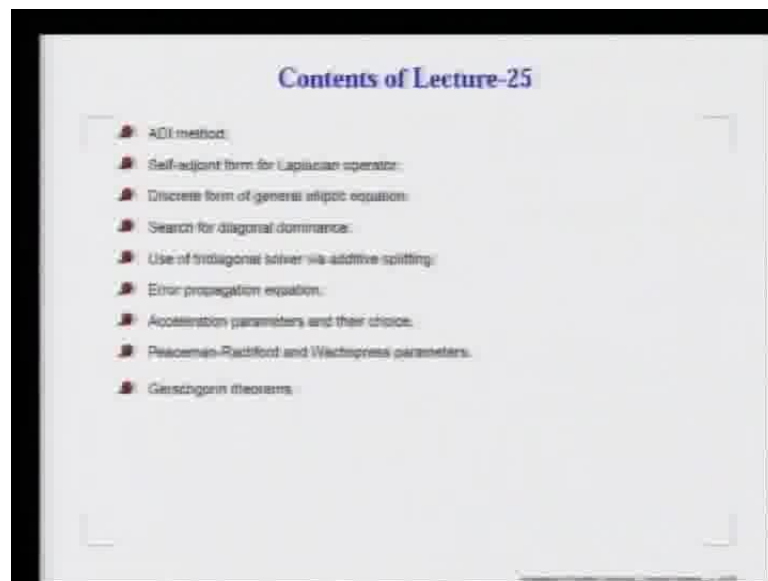
**Indian Institute of Technology, Kanpur**

**Module No. # 01**

**Lecture No. # 25**

In today's lecture number 25, we start our discussion on alternating direction implicit method. We talk about a general elliptic equation, which could represent either the Laplace's equation or the Poisson equation or Helmholtz equation.

(Refer Slide Time: 00:22)

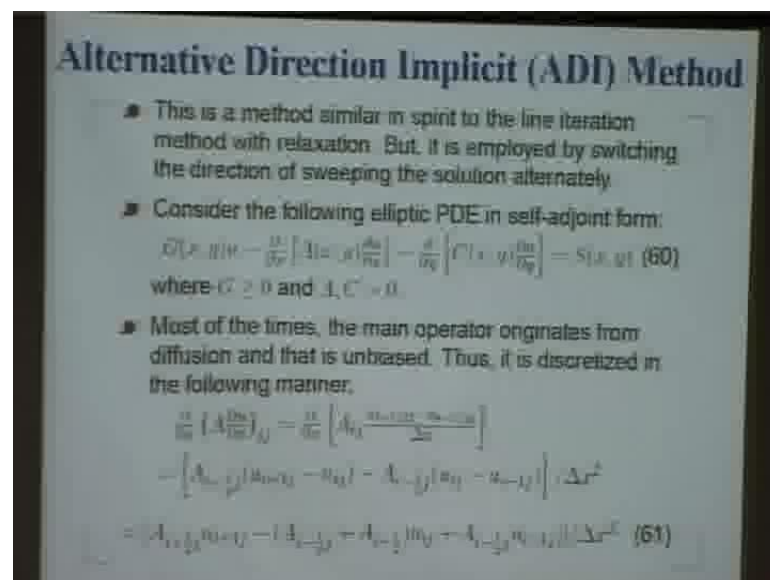


We notice that it requires discretizing the Laplacian operator and we will identify that this Laplacian operator is best discretized by casting it in a self-adjoint form. Once we have this discrete form of general elliptic equation, we show that the self-adjoint form actually leads us to a diagonal e dominant matrix. We also notice that this ADI method actually utilizes the analytic solutions of tri-diagonal linear algebraic equation; that is what we actually do. In a two-dimensional problem, we split it in additive fashion and in each step of the ADI method, we go along one particular direction and since the

directions are alternated; that is the reason we have called this method as alternating direction implicit method.

As we discussed, it is about error propagation properties and we find out that this in ADI method. What we actually obtain is, introduction of acceleration parameters to further accelerate this alternating direction approach. These acceleration parameters are of two kinds: one set was given by Peaceman and Rachford and the other set was given by Wachspress and these acceleration parameters are found out with the help of the pair of theorems due to Gerschgorin. That is what we are going to talk about today.

(Refer Slide Time: 03:02)

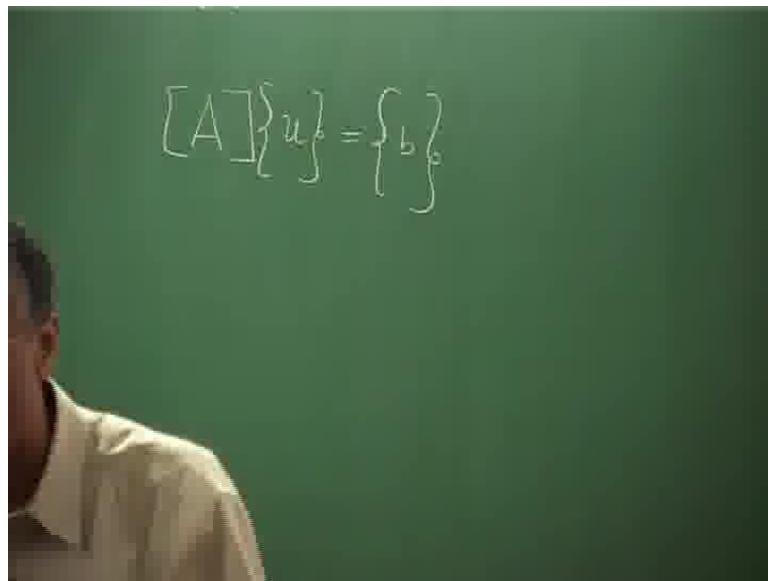


Get back to our discussion on ADI method. See, we make this observation that the ADI method is similar in spirit to line iteration method. We have seen line iteration method is basically the method of choice till 1950s, because they were showing a progression in terms of computational efficiency, speed and benefits; so people were doing line iteration.

However, it was also noted that the strategy by which we go along in choosing these lines are favored boundary conditions from one side or the other. That is why this idea came upon that if we keep switching this direction of solution then what will happen is, we could in an unbiased manner bring the boundary information from all sides. That is desirable for a boundary value problem, so that was the whole idea.

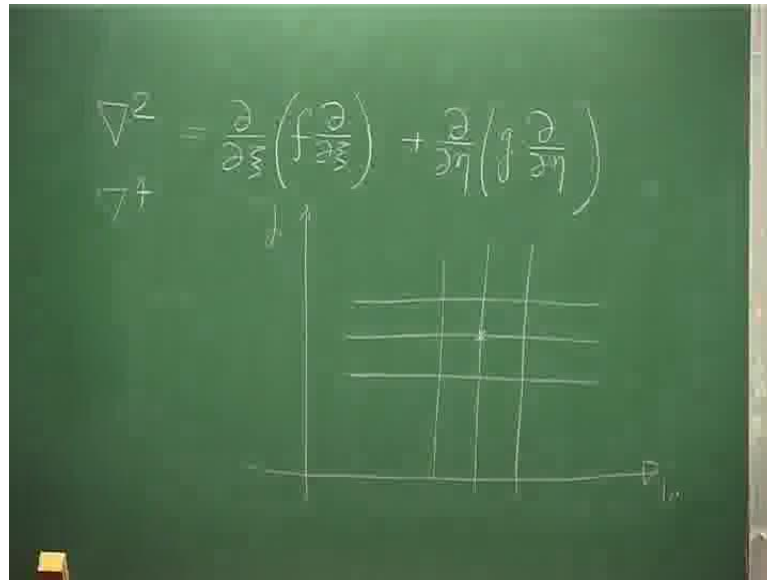
The second thing is of course, I noted to you that most of the time that elliptic differential equation - partial differential equation - that we solved involves Laplacian operator or the bi-harmonic operator. Wherever they appear, they of course, appear as a power of 2. That is the reason that you get complex conjugate as characteristics. What you note that representing this Laplacian or the bi-harmonic operator, you like to do what I have noted down here as a self-adjoint form; we like to have something in a particular fashion that gives certain numerical properties.

(Refer Slide Time: 04:50)


$$[A]\{u\} = \{b\}$$

What are these numerical properties? Some of which we discussed in the last class itself, that we want the associated linear algebraic equation of the following form and we should have some nice property for this matrix A and one of the nicest property that you can think of is the matrix should be symmetric; it would be positive definite, that is what we talked about.

(Refer Slide Time: 05:36)



The chalkboard displays the following equation and diagram:

$$\nabla^2 u = \frac{\partial}{\partial \xi} \left( f \frac{\partial u}{\partial \xi} \right) + \frac{\partial}{\partial \eta} \left( g \frac{\partial u}{\partial \eta} \right)$$

Below the equation is a coordinate system with a grid. The horizontal axis is labeled  $\xi$  and the vertical axis is labeled  $\eta$ . A grid of lines is drawn, with a central point marked by a dot. The grid consists of 3 horizontal lines and 3 vertical lines, creating a 2x2 grid of squares.

To at least achieve symmetry, this form of the PDE is very desirable because for example, let us investigate one of the points - constituent point I have written here, see what we have done here? We have written it down in a form like this (Refer Slide Time: 05:32); this is what we always would like to do.

So, if I have the Laplacian operator of this kind, I would rather like to write it in this form. This is something like your self-adjoint form; we talking about here. See, this A is like your f and C is like your g, why do we do that? Its very operant when you start looking at the discretization of that term, this term here has been discretized in this manner.

How it has been discretized? Let us take a look. Let us say, this is one of the coordinate direction; this is the other one. Let us look at a segment of a point, so this is my i j th point that is where we want to discretize the term. Now, what we try to do is, we try to have a system of discretized equation which remains unbiased. So, if I am trying to discretize in the x direction, I should not have any asymmetric coefficient in the discrete form; that is what we mean by unbiased nature.

How do I do that? We have already seen that if I do a kind of a central differencing, here in this part this  $\frac{\partial u}{\partial x}$  has been written down in this form. What has been done? We have taken the point half a node to the left and half a node to the right divided by the

spacing between these two nodes; that is why this is your equivalent CD2 formulation isn't it; see, we have taken half a node to the left hand to the right.

Of course, we are focusing our attention on the  $i, j$  th node, so we have kept  $a$  here at the  $i, j$  th node. Basically, what we have done in the first step? We have invoked this two points, so this is my  $i$  th node and this my  $j$  th line. In the first step, I am discretizing  $\Delta u \Delta x$  by taking this point and that point; that is what we have done.

Well, you can proceed further. Now, this is the function and that you are differentiating, again you do the same thing. Let me take this first set of point  $A_{i,j}$  times  $u_{i+\frac{1}{2},j}$ . If I do that then, what I am going to do? I am going to take this one shifted half to the right that is why I have written  $A_{i+\frac{1}{2},j}$  and then that is multiplied with this quantity  $u_{i+\frac{1}{2},j}$  that has to be differentiated and if I take a half a point shifted to the right that will give us  $u_{i+1,j}$  and if I take half a point to the left that will give me  $u_{i,j}$ .

This part is coming from the first product and the next part comes from the second product. Now, this  $A$  is now shifted half a point to the left, so that is why we have written  $A_{i-\frac{1}{2},j}$  and of course,  $u_{i-\frac{1}{2},j}$  if I differentiate, I will get half a point to the right, so that will be  $u_{i,j}$  and half a point to the left would give me  $u_{i-1,j}$ . Of course, there was  $1 \Delta x$  and this differentiation would bring in another  $\Delta x$ , so I have the whole thing divided by  $\Delta x^2$ .

So, this is the point but we write it like this (Refer Slide Time: 09:45). You can very clearly notice that discretization of this term invokes the unknown at three points, which are those: the central point itself  $u_{i,j}$  then, one point to the right  $u_{i+1,j}$  and this is this.

(Refer Slide Time: 10:22)

**Alternative Direction Implicit Method (cont.)**

• If we use the short-hand notation:  $h = \Delta x$  and  $k = \Delta y$ , then

$$-\frac{h^2}{12} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = -\frac{h^2}{12} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = -\frac{h^2}{12} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} \quad (10)$$

$$\text{and } -\frac{h^2}{12} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = -\frac{h^2}{12} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = -\frac{h^2}{12} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} \quad (11)$$

where  $\frac{\partial^3 u}{\partial x^3} = \frac{1}{h^3} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = \frac{1}{h^3} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = \frac{1}{h^3} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n}$

and  $\frac{\partial^3 u}{\partial x^3} = \frac{1}{h^3} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = \frac{1}{h^3} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n} = \frac{1}{h^3} \left[ \frac{\partial^3 u}{\partial x^3} \right]_{m,n}$

$$M = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & 0 & \dots & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \dots & 0 \\ 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & \dots & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \quad \begin{Bmatrix} u_{1,n} \\ u_{2,n} \\ u_{3,n} \\ \vdots \\ u_{N-1,n} \end{Bmatrix} = \begin{Bmatrix} \frac{1}{6} \\ \frac{1}{6} \\ \frac{1}{6} \\ \vdots \\ \frac{1}{6} \end{Bmatrix}$$

Now, what you notice though that there is the form appearing from this discretization; this is what we have try to explore here. Let us take for the sake of simplification of notation describe  $h$  as  $\Delta x$  and  $k$  as  $\Delta y$  then, the quantity that we were discretizing was  $\Delta x$  of  $A \Delta u \Delta x$ . If I multiply that quantity by  $h k$  and I purposely multiplied it with the minus sign, so that the diagonal term appears with a plus sign and the half diagonal terms appear with a negative sign - you recall - we enunciated some property of  $A$  matrix property to be property  $A$  we called and that was desired that your diagonal elements are positive half diagonal elements are negative. Then, they would also have the diagonal dominance property to get a good iterative solution strategy.

So, if I write it like this; so if I write it there at  $m$  nth node, I get this structure and it is very easy for you to clearly see the tri-diagonal matrix appearing here. From the first term which I called as  $HU$ ; so  $HU$  here is written in this form.

Basically, we keep writing it for different values of  $m$  starting from the first interior point. So, if I look at the boundary point as  $u_{0,n}$  then, I will start this discretization from  $u_{1,n}$  and when I do that, I will see  $u_{1,n}$  would be the my **beginning** starting post and that would have a point  $u_{2,n}$  and  $u_{0,n}$ ;  $u_{0,n}$  is the boundary point, so that would not stay on this side; that goes to the right hand side, this is a known quantity.

When I write in terms of the unknown then, the first line gives me two entries that will have the entry that would be here  $2b_{mn}$  and this quantity would be on the super diagonal that is the minus  $a_{mn}$ . Now, of course, once you move away from the near boundary point you start having all these three points. For example, 0.2 onwards you will get 1, 2 and 3 and then, you start getting this regular tri-diagonal structure, so this is what it is.

Now, what is this  $a_{mn}$ ? Remember, this was divided by  $h^2$ ; so when I multiply by  $h_k$  that would give me a factor of  $k$  by  $h$ . So, I have this entries given as a constant multiplier  $k$  by  $h$  in all three and what you also notice a particular nice property that the half diagonal terms, if I take their magnitude and add it up that is what gives me the diagonal quantity, so  $2b_{mn}$  is nothing but, equal to  $a_{mn}$  plus  $c_{mn}$ .

Why do we do that? Well, you recall that was one of our requirements for diagonal dominance. We wanted that at the most the diagonal term in magnitude should balance the sum of all the half diagonal term and that is what is guaranteed by this kind of discretization, because this central entry is always balance by this. Where do you get strict inequality? You do get the strict inequality on the first row and the last row. So, by doing this kind of discretization, what you are ensuring is that you would have a very nice diagonal dominance property of the system.

Now, you look at the other term that we had something like this, there we had written like  $\Delta y$  into  $c \Delta u \Delta y$  again, you multiply by minus  $h_k$  and you will get in terms of this coefficients  $\alpha$ ,  $\beta$ s and  $\gamma$ s and they also have that same kind of property. Now, you notice that here the constant multipliers are  $h$  by  $k$ ; that same property is shared here the diagonal entry in magnitude is some of the magnitude of the half diagonal entries.

(Refer Slide Time: 16:32)

**Alternative Direction Implicit Method (cont.)**

- Finite difference analogue is:  $(H + V)U = K$  (64)
- The 'G'-term adds to the diagonal dominance and in convergence improvement.
- Let us instead investigate:  $(H + V)U = K$  (65)
- For any matrices 'D' and 'E', Eqn. (65) is equivalent to:

**NON-SINGULAR**  $\begin{cases} (H + D)U = K - (V - D)U & (66a) \\ (V + E)U = K - (H - E)U & (66b) \end{cases}$

- Peaceman & Rachford used:  $D = E = \rho_0 I$  so that I h.s. of

$$(H + \rho_0 I)U^{(k+1/2)} = K - (V - \rho_0 I)U^{(k)} \quad (67a)$$

$$(V + \rho_0 I)U^{(k+1)} = K - (H - \rho_0 I)U^{(k+1/2)} \quad (67b)$$

are easy to invert TRIDIAGONAL MATRICES.

So, if you now go back and take a look at the equation 60, we have just now talked about discretization of this part. In addition we all have a term like this G into u where do we get it, those of you are familiar with little bit of electrostatics; you would know that is the kind of term you get in Helmholtz equation.

You would get this kind of term there, so that will be like if we are discretizing at the point ij so this will be just simply Gij times u ij and this is you have some kind of a forcing that you get in Poisson equation. Poisson equation you get del square f equal to G or something, so that right hand side the function S is that term that accounts for this.

So, in a generic sense this equation 60 that we have in front of us is a potpourri of everything. We have the Helmholtz equation as a special case; Poisson equation is a special case, you can also treat it has a Laplace's equation. Laplace's equation of course, G and S will be equal to 0, so this is what we are talking about.

Now, having discretized those second derivative terms in the self-adjoint form we get this as the finite difference analog. We had H times U giving us the first set of term del del x of a del U del x term. So that is the H it is their then, this is the second term and the term I told you which makes it a keen to the Helmholtz equation G times U adds on to what I have written there is a capital sigma K H times G omega, I have multiplied the whole equation by K H that is what we have done.



If you recall, when I wrote in the previous slide I did purposely multiplied all the terms by  $H$  and  $K$ , so that certain symmetric nature showed up. What happens is that  $\sum U$  that we are writing here is basically  $H K$  times  $G$  and forcing term on the right hand side will be  $H K$  times  $S$  evaluated at the point in question  $m n$ .

Now, once I write this discrete equation 64; you can place where each of this term go. For example the  $G$  term, where does it go? It goes to the diagonal because we are looking at the  $m n$ th point that goes along the diagonal. So,  $\sum U$  would actually add on to the diagonal entry; so that would actually help the diagonal dominance property. If I have without the  $G$  term already we have shown that diagonal dominance property is there then, in addition to that if I add this  $G$  term this will strengthen the diagonal dominance property and we have already noted that such diagonal dominance adds in convergence.

Let us make our job little more-tougher, so let us drop the  $G$  term;  $G$  term is easy to make - I mean - it automatically makes your life easier. Let us try to investigate a case, where is somewhat little more challenging, we drop that  $G$  term and instead we start looking at this equation 65. Basically it is nothing but, an attempt to solve a equivalent Poisson equation we are solving a Poisson equation here that is given by 65.

This is the basic thing that we have done so what is new, apart from the discretization what is new? Well, this is where the people working in the area ADI came up with the idea that we could split this equation  $H$  plus  $V$  operating on  $U$  to provide that right hand side  $K$  in 2 steps.

(Refer Slide Time: 20:10)

$$[A]\{u\} = \{b\}$$

$$-\frac{\partial}{\partial x} \left( A \frac{\partial u}{\partial x} \right) - \frac{\partial}{\partial y} \left( C \frac{\partial u}{\partial y} \right) = K$$

$$(D + H)U = - VU + K + DU$$

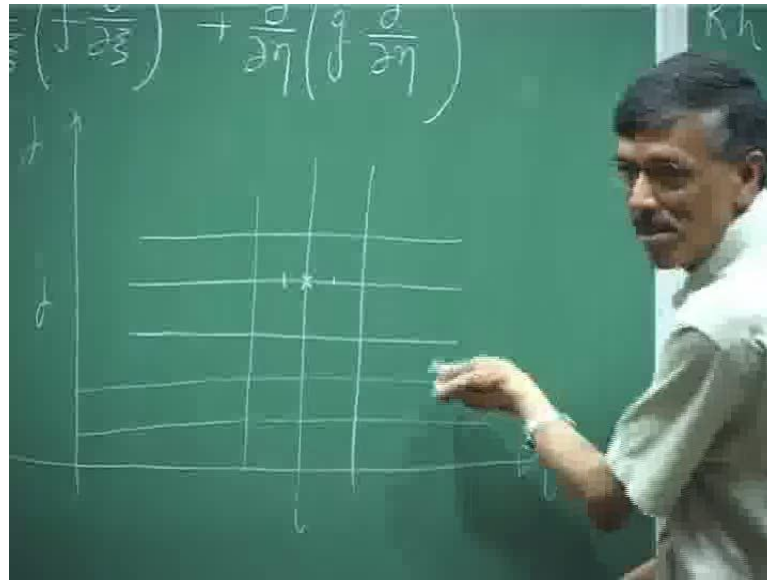
What we do actually, is given in equation 66a and 66b. In the first half step I add DU term on both sides then what happens on the left hand side I have H plus D into U and right hand side K was of course, there VU has been putted on the right hand side but, since I have added a plus D U, I again here add a plus D U here note the minus sign together will make it.

We have just simply added the same term on both sides. What is the structure of H plus D matrix? It is still a tri-diagonal matrix, H matrix itself had some bare bone diagonal dominance property to that we are adding something as D. So, if D and E happens to be diagonal matrix then we are adding to a diagonal dominance then what we are basically talking about, 66a is what? It is a line iterative method, isn't it?

What is your H operation? Is this derivative; basically, we are looking at this equation now (Refer Slide Time: 21:14) this is what we are trying to solve. From here in a discrete form I get HU and from here we get discrete form VU, so that is what we have written there 65 is H plus VU.

Now, what I am suggesting to you that put this term in the right hand side. So, what will happen? Then I will get this, so this itself is a going to give me a tri-diagonal form so that I am adding it entry which announces the diagonal dominance on the left hand side.

(Refer Slide Time: 22:40)



What we end up doing actually then we are basically writing it exactly like what I have written there in 66a. What happens is I can now solve that 66a along the  $j$  equal to constant lines that is what will amount to. Basically, what we are suggesting that we will be solving line by line in this fashion and when I solve that; that would be like solving  $HU$  is equal to  $K$  minus  $VU$  that is what we do in line iterative method, isn't it?

We put the  $y$  derivative on the right hand side and we keep solving along these lines horizontal lines and much that is what our line iterative method has been design to do.

(Refer Slide Time: 23:16)

### Alternative Direction Implicit Method (cont.)

- Finite difference analogue is:  $(H - V)U = \sum_{j \neq i} G_{ij} U_j - \sum_{j \neq i} G_{ji} U_j$  (64)
- The ' $G$ '-term adds to the diagonal dominance and in convergence improvement.
- Let us instead investigate  $(H + E)U = K$  (65)
- For any matrices ' $D$ ' and ' $E$ ', Eqn. (65) is equivalent to:
 

**NON-SINGULAR**

{

$(H + DE)U = K - (V - D)U$  (66a)

{

$(V - E)U = K - (H - E)U$  (66b)
- Peaceman & Rachford used:  $D = E = \rho_0 I$  so that I h s, of
 

$(H + \rho_0 I)(U^{k+1/2}) = K - (V - \rho_0 I)U^{(k)}$  (67a)

$(V - \rho_0 I)U^{(k+1)} = K - (H - \rho_0 I)U^{(k+1/2)}$  (67b)

are easy to invert TRIDIAGONAL MATRICES.

Now, what we are doing? Making that line iterative method little stronger and that is why we are adding this  $D$  matrix, so that is what we are doing in this thing. Now, what happens is of course, if  $H$  matrix was little hyper sensitive that it had not a very well condition property; by adding the  $D$ , we have ensured that we would not have problem in solving this equation. So, because  $H$  plus  $D$  would be a well condition matrix; it can be inverted in a mathematical sense and computational sense, so that is why I am calling it as non-singular. That is the first step, I keep solving it like this, and I go from line to line like this start it up.

Now, in the next half step what I do is, I change my direction. Now, I would be solving it in this fashion that is what that second equation implies that  $V$  plus  $E$  into  $U$  is this. So, it basically the same thing, we are doing - treating. Now, what we could do is, we could put some kind of iteration index on this so that is what I have done.

What I have done is written it down here; what Peaceman and Rachford suggested that you purposely take this  $D$  and  $E$  matrix as strictly diagonal matrix. How do you do that? Well, you just define  $D$  and  $E$  is equal to some acceleration parameter, we will talk about this acceleration parameter shortly, which has given the notation here  $\rho$   $n$  times the identity matrix then, I say we are actually doing this.

So, if I have a solution at the  $k$  th level solving 66a is equivalent to solving this. So, I get a half step solution which I am calling it as  $k$  plus half. Having obtain that  $k$  plus half solution then, I go to the next half cycle that is the my forcing term and I solved this equation and that completes a full cycle.

First, I do a vertical sweep followed by a horizontal sweep and that constitute one operation of ADI method. Why did we do that? Well, that is what I have been suggesting to you that look the tri-diagonal matrix plays a very central role in computing and these two matrices  $H$  plus  $\rho$   $KI$  and  $V$  plus  $\rho$   $KI$  they are both tri-diagonal matrices.

Now, if we can choose this  $\rho$   $K$  with lot of care then, we could not only ensure diagonal dominance; we can also control error better and that is what we are going to study now. How the choice of this  $D$  and  $E$  matrix actually influences the success of the method.

(Refer Slide Time: 27:02)

### ADI Method (cont.)

- At convergence:
 
$$(H + \rho_k I)U^{(k+1)} = b - (V - \rho_k I)U^{(k)} \quad (68a)$$

$$(V + \rho_k I)U^{(k+1)} = b - (H - \rho_k I)U^{(k)} \quad (68b)$$
- Denoting  $U^{(k+1/2)} = U^{(k+1)} = U^{(k+1/2)}$  we get from Eqns. (67) and (68):
 
$$(H + \rho_k I)U^{(k+1/2)} = -(V - \rho_k I)U^{(k)} \quad (70a)$$

$$(V - \rho_k I)U^{(k+1/2)} = -(H - \rho_k I)U^{(k+1/2)} \quad (70b)$$
- Thus we obtain:
 
$$U^{(k+1)} = -(V - \rho_k I)^{-1}(H - \rho_k I)U^{(k+1/2)} \quad (70a)$$

$$= (V - \rho_k I)^{-1}(H - \rho_k I)(H + \rho_k I)^{-1}(V - \rho_k I)U^{(k)} \quad (70b)$$

$$= T^{(k)}U^{(k)}$$
- $T^{(k)}$  is the Error reduction matrix and
 
$$\|U^{(k+1)}\| \leq \|T^{(k)}\| \|U^{(k)}\|$$

So, if I have chosen those half steps and I have rewritten those equations but, now I am saying look, I have performed an infinite number of such operations. Once, I have done that many operations then, I do not write it  $K$  plus half  $K$  everything is at the convergence they have become indistinguishable what you have on the left hand side and right hand side. So, this is what we get 68a and 68b are the status of the system at convergence.

(Refer Slide Time: 27:47)

### ADI Method (cont.)

- At convergence:
 
$$(H + \rho_k I)U^{(k+1)} = b - (V - \rho_k I)U^{(k)} \quad (68a)$$

$$(V + \rho_k I)U^{(k+1)} = b - (H - \rho_k I)U^{(k)} \quad (68b)$$
- Denoting  $U^{(k+1/2)} = U^{(k+1)} = U^{(k+1/2)}$  we get from Eqns. (67) and (68):
 
$$\rightarrow (H + \rho_k I)U^{(k+1/2)} = -(V - \rho_k I)U^{(k)} \quad (70a)$$

$$(V - \rho_k I)U^{(k+1/2)} = -(H - \rho_k I)U^{(k+1/2)} \quad (70b)$$
- Thus we obtain:
 
$$U^{(k+1)} = -(V - \rho_k I)^{-1}(H - \rho_k I)U^{(k+1/2)} \quad (70a)$$

$$= (V - \rho_k I)^{-1}(H - \rho_k I)(H + \rho_k I)^{-1}(V - \rho_k I)U^{(k)} \quad (70b)$$

$$= T^{(k)}U^{(k)}$$
- $T^{(k)}$  is the Error reduction matrix and
 
$$\|U^{(k+1)}\| \leq \|T^{(k)}\| \|U^{(k)}\|$$

Now, I can denote an error vector which I have shown here for the  $K$  plus half step that will be the correct solution minus the current solution. If I do that then, what I can do is I can subtract these two equations from my commutative equations in 67 and subtraction would give me this equation,  $H$  plus  $\rho K I$  into  $E$  and this (Refer Slide Time: 27:46). You can see in both the equations  $K$  is common so as far as error is concerned the  $K$  cancels out.

The error evolution from  $k$  th step to  $k$  plus half step is entirely determined by the structure of  $H$  and  $V$  matrix and by the way, we have added this acceleration parameter  $\rho K$ . In the second half, we go from  $K$  plus half and we arrive at  $K$  plus 1 step so that is what my error is. So, if I combine the both the steps then, what do I get  $e_{k+1/2}$  would be  $V$  plus  $\rho K I$  inverse operating on this and there is this minus sign setting up front operating on  $e_{k+1/2}$  but, from the first equation, I can get  $e_{k+1/2}$  is nothing but, minus of  $H$  plus  $\rho K I$  inverse into this, so that is precisely what we have done.

What we seen that one step of ADI involves the error to migrate from  $e_k$  to  $e_{k+1}$  through this matrix  $T$  of  $k$  which I have called it has the error reduction matrix. You can very clearly see what this  $T$  of  $k$  is;  $T$  of  $k$  is this whole product set of these four matrices. Please do understand that when I write this inverse it is not such a great difficult task to undertake because these are tri-diagonal matrix; we have exact solution by Thomas algorithm, so that is not a question that we should be bothered about.

So, once I see the error at successive steps are related by the error reduction matrix then, I can form a kind of a norm for measuring them. So, that is what is written here that the error at the  $k$  plus 1 th level is bounded by what was my error of the previous step times the norm of the error reduction matrix.

(Refer Slide Time: 30:34)

### Alternative Direction Implicit Method (cont.)

- Finite difference analogue is:  $(H + V + \sum_{h \in E_{\text{int}}} U) = K$  (64)
- The 'G'-term adds to the **diagonal dominance** and in convergence improvement.
- Let us instead investigate  $(H + V)U = K$  (65)
- For any matrices 'D' and 'E', Eqn. (65) is equivalent to:
 

**NON-SINGULAR**

$$\begin{cases} (H + D)U = K - (V - D)U & (66a) \\ (V + E)U = K - (H - E)U & (66b) \end{cases}$$
- Peaceman & Rachford used:  $D = E = \rho_n I$  so that l.h.s. of
 

$$(H + \rho_n I)U^{(k+1/2)} = K - (V - \rho_n I)U^{(k)} \quad (67a)$$

$$(V + \rho_n I)U^{(k+1)} = K - (H - \rho_n I)U^{(k+1/2)} \quad (67b)$$

are easy to invert **TRIDIAGONAL MATRICES**.

There are various ways of defining the norm but, we have already seen that we do it in terms of the spectral radius. We have done that but, none the less will still figure out certain interesting properties of this. Please take a look at the previous slide that in the definition of this T matrix, the V matrix appears in the beginning and in the end and H matrices are sitting in the middle.

(Refer Slide Time: 30:39)

### ADI Method (cont.)

- As similarity transformation does not affect spectral radius,
 
$$\begin{aligned} \bar{T} &= (V - \rho_n I)U(V - \rho_n I)^{-1} \\ &= (V - \rho_n I) \left\{ (V - \rho_n I)^{-1} [H - \rho_n I] (V - \rho_n I)^{-1} \right\} \\ &= (V - \rho_n I) \left\{ (V - \rho_n I)^{-1} \right\} \\ &= (H - \rho_n I)(V - \rho_n I)^{-1} (V - \rho_n I) \end{aligned}$$
- The matrices in the square brackets have real eigenvalues.
- The eigenvalues of  $\bar{T}$  are of the form:
 
$$\left( \frac{\mu - \rho_n}{\mu + \rho_n} \right) \left( \frac{\nu - \rho_n}{\nu + \rho_n} \right)$$

where,  $\mu, \nu$  are the eigenvalues of  $H, V$  respectively

So that is little unwieldy; that unwieldy structure can be rectified by performing a similarity transformation. One of the properties of similarity transformation is that if I

have a matrix, I perform a similarity transformation Eigen values, Eigen vectors do not change. We have to take it from it, I am not going to say more than that. This is a well established result that if I perform a similarity transformation, how do I perform a similarity transformation? I take the original matrix  $T$  that we have here and then, I pre multiply by the matrix and post multiply by its inverse, so that is what we have done. We have actually done here that  $V + \rho k I$  multiplying  $T$  and  $V + \rho k I$  inverse post multiplying that.

Then, what happen? I can put in the expression for the  $T$  matrix shown in the braces and then, I get this. Now, actually what has happened? I have a sort of it ((there)) of product term first of which is totally dependent on  $H$  matrix; the second is dependent on  $V$  matrix.

So, this  $T$  tilde matrix that we have obtained is determined in terms of this particular product form. Now, if I look at the tri-diagonal matrix with that property that we have discussed; we can show that the Eigen values of those matrices will be real. The entries are real plus those properties that we have enunciated will ensure that they will have real Eigen values.

So, if I have Eigen values of  $H$  and  $V$  matrix as  $\mu$  and  $\nu$  then, I can see that the Eigen value of this  $T$  tilde matrix would be given by this; so instead of  $H$ , I write  $\mu$ . So, this first factor is  $\mu - \rho$  and this is inverse, so that will go in the denominator  $\mu + \rho$ . The second factor will give me  $\mu - \rho$  and this inverse will go down stairs to give me  $\nu + \rho$  in the last expression.

Why have we written like this? It is very clearly shown irrespective of whatever may be the value of  $\mu$  and  $\nu$  to be each of these products within the first bracket are guaranteed to be less than 1, can you see that? That whatever may be the value of  $\mu$  and  $\nu$  this factor as well as this factor is always less than 1.

Can you see the tremendous benefits that we have derived out of this particular operation. Whatever may be the spectral radius of  $H$  and  $V$  matrix, they could be quite not so well behaved; they could be very close to 1, so that it will struggle like what we have seen.



Even then these factors or each one of them is always going to be less than 1. Basically what does it imply? That is understood here that my error in successive iteration is moderated by their norm of the T matrix and that norm of the T matrix we have just now ensured is guaranteed to be less than 1, so this shows a nice convergent iteration.

(Refer Slide Time: 35:04)

### ADI Method (cont.)

- As similarity transformation does not affect spectral radius,  

$$\tilde{T}(p_k) = (V - p_k I) T (V - p_k I)^{-1}$$

$$= (V - p_k I) \left\{ (V - p_k I)^{-1} (H - p_k I) (H - p_k I)^{-1} \right. \\ \left. (V - p_k I) (V - p_k I)^{-1} \right\}$$

$$= (H - p_k I) (H - p_k I)^{-1} (V - p_k I) (V - p_k I)^{-1}$$
- The matrices in the square brackets have real eigenvalues.
- The eigenvalues of  $\tilde{T}$  are of the form:  

$$\left( \frac{\mu - p}{\mu + p} \right) \left( \frac{\nu - p}{\nu + p} \right)$$
 where,  $\mu, \nu$  are the eigenvalues of  $H, V$  respectively.

(Refer Slide Time: 35:12)

### ADI Method (cont.)

- After  $T$  iterations:  $x^{(k+1)} = \prod_{i=1}^T \tilde{T}(p_i, \mu^{(k)})$
- The eigenvalues of  $\prod_{i=1}^T \tilde{T}(p_i)$  are given as:  

$$\lambda(p, \mu) = \prod_{i=1}^T \frac{(\mu - p_i)(\mu - p_i)}{(\mu + p_i)(\mu + p_i)}$$
- Thus,  

$$\left\| \prod_{i=1}^T \tilde{T}(p_i) \right\| = \lambda \left( \prod_{i=1}^T \tilde{T}(p_i) \right) \leq \prod_{\mu \in \sigma} \left\| \frac{(\mu - p_i)(\mu - p_i)}{(\mu + p_i)(\mu + p_i)} \right\|$$
- One practically estimates upper & lower bounds of the eigenvalues from  $a \leq \mu \leq b$  and  $u \leq p \leq \beta$

So, in doing that what we need to do is of course, find out the Eigen values of H and V matrix and we have to also find out what this rho is, that acceleration parameter that we

talked about. Now, look at the way that we have written down the linear algebraic equation that  $e_{k+1}$  is related to  $e_k$  by the multiplication of  $T$ .

Now, if I do  $l$  such iterations, so I go from  $e_k$  to  $e_{k+l}$  then, I find it is nothing but, taking a product of this successive  $T$  matrices. This is where we also note that earlier also I had added a subscript with  $\rho$ , here also I have added a subscript with  $\rho$  here  $i$  it means that from on each iteration I could take this  $\rho$  to be different.

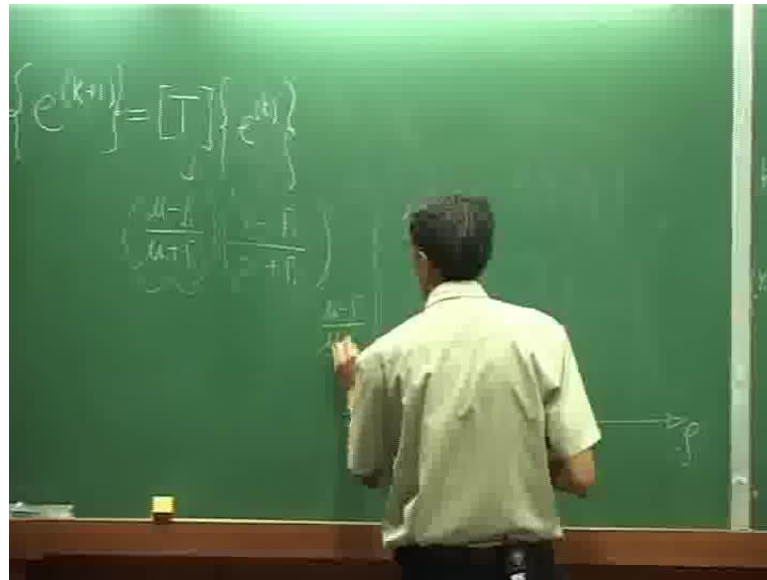
How do we choose it, we will come to that but, there it shows that we have added degree of freedom in choosing the acceleration parameter  $\rho_i$ . So such that the error after  $l$ th iteration is determined by the Eigen value of this product matrix of  $T$ .

So, I would have  $T$  of  $\rho_1$ ,  $T$  of  $\rho_2$  and so on so forth up to  $T$  of  $l$  that is what this product operation is signifying. So, I call the Eigen values of that product in terms of  $\lambda$ . Now,  $\lambda$  is a function of those Eigen values of the constituent  $H$  and  $V$  matrix so  $\mu$  and  $\nu$ ; they will look like this and they will also depend on the acceleration parameter  $\rho_i$  - I choose - so  $i$  go from 1 to  $l$ .

So, if I look at the norm of the spectral radius of that  $\lambda$  matrix that would be given here, let me call that as capital  $\lambda$  matrix. So that would be nothing but, take this product form and look for its maximum value, when  $\mu$  and  $\nu$  take its successive values. How many values of  $\mu$  and  $\nu$  will get? That will be determined by the dimension of  $H$  and  $V$  matrix. If I have an  $H$  matrix which is dimension is  $m$  by  $n$  then, what do I get? I will at most get  $n$  distinct Eigen values, so that is precisely what we are doing.

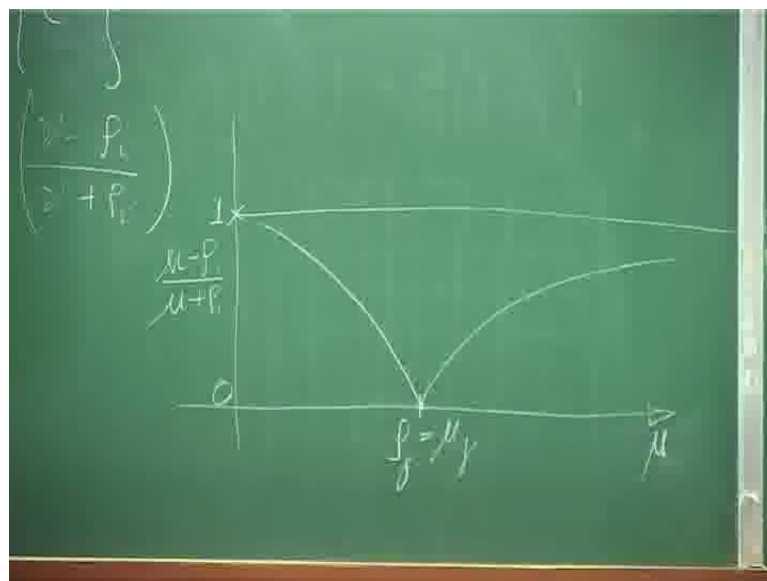
We are scanning through all those  $m$  and  $n$  Eigen values of  $\mu$  and  $\nu$  and trying to find out what is the maximum of it because, we know when we adopt this strategy eventually what matters is the spectral radius of this product matrix.

(Refer Slide Time: 38:25)



Now, let me draw a diagram to tell you what we get out of this operation by this T matrix. See, what we have written down just now is that  $e^{k+1}$  is given by the T matrix, so this is a vector and this is operating on this. Now, what we found is that this itself in a non fashion gives me like  $\mu - \rho_i$ ,  $\mu + \rho_i$  and  $\nu - \rho_i$  and  $\nu + \rho_i$ .

(Refer Slide Time: 39:40)



Let us look at one of this factor, let me plot this. On this side you plot it as a function of  $\rho$  and on this side I am plotting  $\mu - \rho_i$  by  $\mu + \rho_i$ . Now, if I choose  $\rho_i$

as 1 of the Eigen value of H matrix what will happen, the numerator would be equal to 0 for that Eigen value.

So, if I am choosing a value  $\rho_j$  equal to let us say, call that as  $u_j$  then, what I am going to get for that combination so for different values of, well I think I should plotted  $\mu$  here the whole spectrum of Eigen values that we are looking at. We are trying to find out what this factor does for different spectral component of this Eigen values.

What happens is that value when that acceleration parameter one side with one of the Eigen value that is what this factor is 0. Anything that is less than this, will go like this (Refer Slide Time: 40:32). The facts that I could choose your acceleration parameter coincident with one of the Eigen value and emulates that component of the error because that is where this is 0, where do you get the maximum? See, if  $\mu$  goes to infinity then of course, the value will such way to at the most 1 irrespective of whatever row you choose then, what we was finding out that this is the kind of  $(( ))$  that we get.

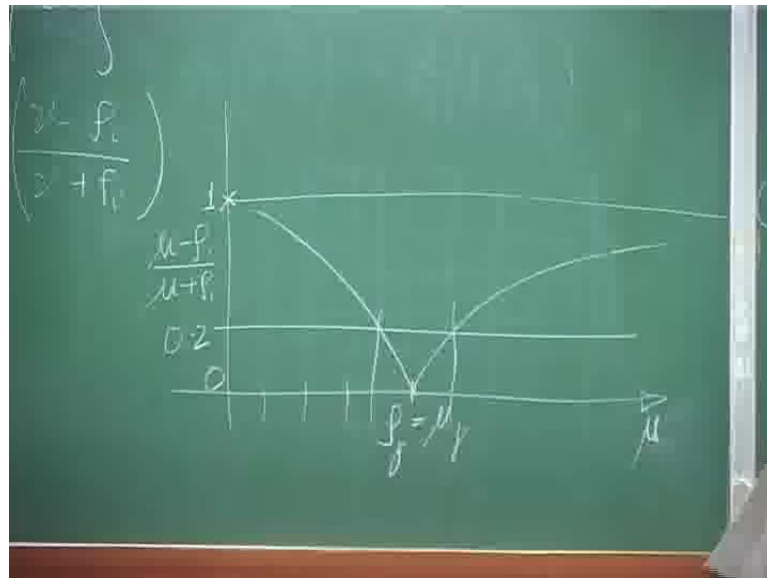
So, all such choice of  $\rho$  acceleration parameter we make actually, brackets the corresponding error component between 0 and 1 and that is exactly what we wanted, that is the whole idea of error reduction. What happens is, then if I have  $\mu_1$  to  $\mu_n$  what I could do is, I could keep choosing my rows as those Eigen values themselves. If I choose  $\rho_1$  equal to  $\mu_1$  then, I know on that iteration any error that is proportional to  $\mu_1$  has been taken care of.

So, this is the whole idea of looking at it and we are also being assured here by this expression that what happens is essentially is the spectral radius of this is what matters most and that is what we want to do but, think of the possibility that if you have taken say 500 point in the x direction 500 point in the y direction T. You have to calculate 500 Eigen values in a for H matrix, 500 Eigen values for V matrix and will be doing this iteration.

First time you will take  $\mu_1$  then, next you take  $\mu_1$  equal to  $\rho_1$  then  $\mu_2$  equal to  $\rho_2$  I mean  $\rho_2$  equal to  $\mu_2$  and so on so forth, but then you are expecting to annulate all the error component; you have to take the full cycle 500 such things and that is a luxury.

For each step of solution method for elliptic equation, if you have to take that many numbers of iteration to take care of one cycle then, you are in ((diastric)) you know, if I have to do that most of the computations will come to halt. What we try to do in our regular day today activity; we tried get it within 5 or 6 iterations, if it is not done then, we know we have a slow method we have to improve upon it.

(Refer Slide Time: 43:48)

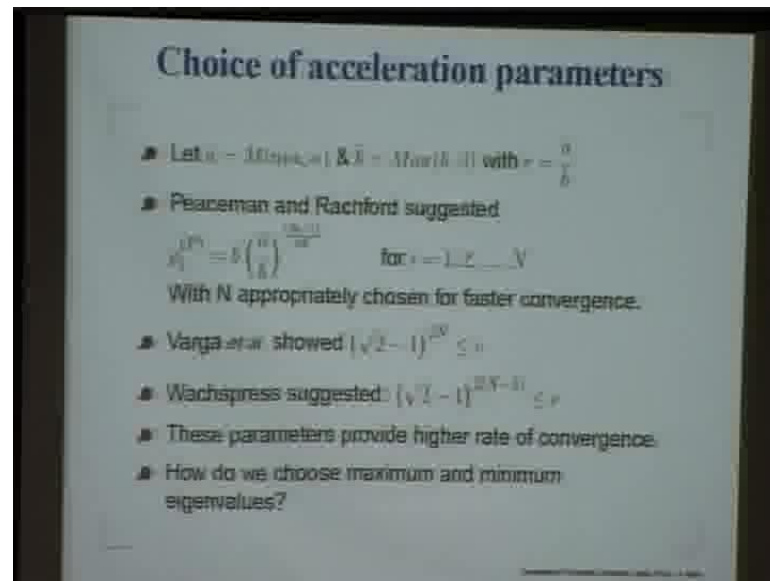


What it actually tell, that even if I know all the Eigen values of H and V matrix it is not very practical to use all those informations up. What you are noticing that when I choose, say  $\rho_j$  equal to  $\mu_j$  I have whole set of range over which this is something like 0.2.

So this range all the error component has been reduced by 80 percent in that step. What I could do is a choice of a selective such points would take care not only that particular point but also its neighborhood. So, I do not need to exhaust all the Eigen values themselves; I could take some selective values and that is what I must take care that practical estimate.

How many such values we should take? How should we go about it? That is guided by this statement that we have the Eigen values whose ranges are given let us say, the  $\mu$  ranges between  $a$  and  $b$  and  $\nu$  ranges between  $\alpha$  and  $\beta$ .

(Refer Slide Time: 45:02)



**Choice of acceleration parameters**

- Let  $\alpha = \lambda(\lambda_1 + \lambda_2)$  &  $\beta = \lambda(\lambda_1 - \lambda_2)$  with  $r = \frac{\alpha}{\beta}$
- Peaceman and Rachford suggested:  
$$x_i^{(p)} = b \left( \frac{\alpha}{\beta} \right)^{\frac{(p-1)(N-i)}{N}} \quad \text{for } i = 1, 2, \dots, N$$
  
With N appropriately chosen for faster convergence.
- Varga et al. showed  $(\sqrt{2} - 1)^{2N} \leq \epsilon$
- Wachspress suggested:  $(\sqrt{2} - 1)^{2(N-1)} \leq \epsilon$
- These parameters provide higher rate of convergence.
- How do we choose maximum and minimum eigenvalues?

Then what we would do is, we will try to find out a sensible number of such acceleration parameters lying between these bands and we will try to work it out. How do we do it? Well, we want to automate. So what we try to do is, we try to find out the minimum of a alpha here and call that as a bar and the maximum of b and beta that we call as b bar we can find out their ratio that is like something like your stiffness ratio.

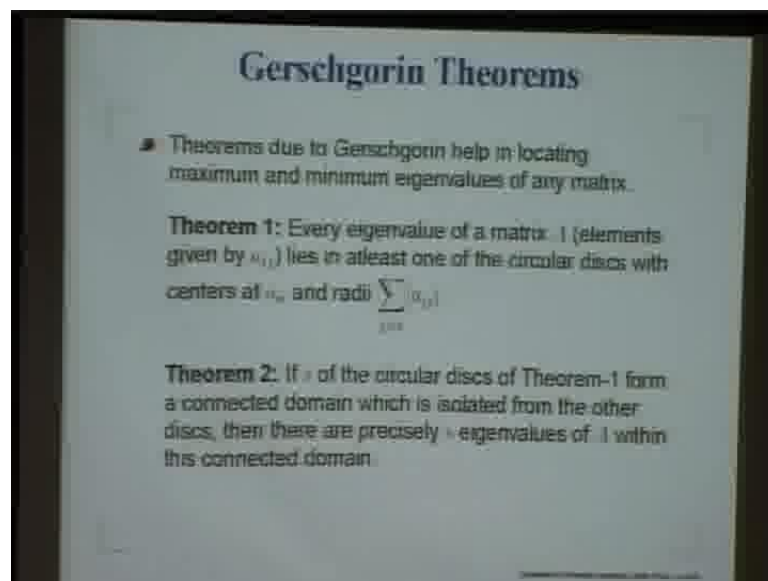
If you recall in the discrete equation, I had that form H by K, K by H so what you find, here also I am talking about how the problem is stiff in the x direction, how the problem is stiff in the y direction that is given by this Eigen value spectrum that is determined by a b and alpha and beta and the ratio of this global maximum and minimum that determines how fast it can be done.

So, Peaceman and Rachford actually suggested that we do it like this. We take a sequence of n of those acceleration parameters indicated with a superscript p to indicated that due to Peaceman and Rachford, so that is given by this. What we are doing actually? You are taking n such acceleration parameters between a bar and b bar that exhaust the whole possibilities; a bar is the lowest value b bar is the highest value, so that is what you do.

Well, there is a bit of mathematics that goes behind all this. How do you choose that n? There are couple of ways Varga and his colleagues have shown that if you know this

ratio  $\bar{a}$  by  $\bar{b}$  that is  $c$ , then you should choose  $n$  in such a way that this inequality satisfies that  $\sqrt{2} - 1$  raised to the power  $2N$  should be less than or equal to  $c$ . There was another person Wachspress; he suggested alternative form and what we get is much higher rate of convergence by adopting this method. Now, what is it that remains to be settled? How do I find out those quantities given in the first line the minimum and maximum of an Eigen value without setting too much?

(Refer Slide Time: 47:23)



We have to compute and we cannot do much with trying to find out your Eigen value spectrum of a 500 by 500 matrix that is too much that it is a problem which may be 1000 times more difficult than your original problem, so you do not want to do that. So, that is where there is couple of theorems due to Gerschgorin that comes very handy. This is very practical and very useful, let us use it.

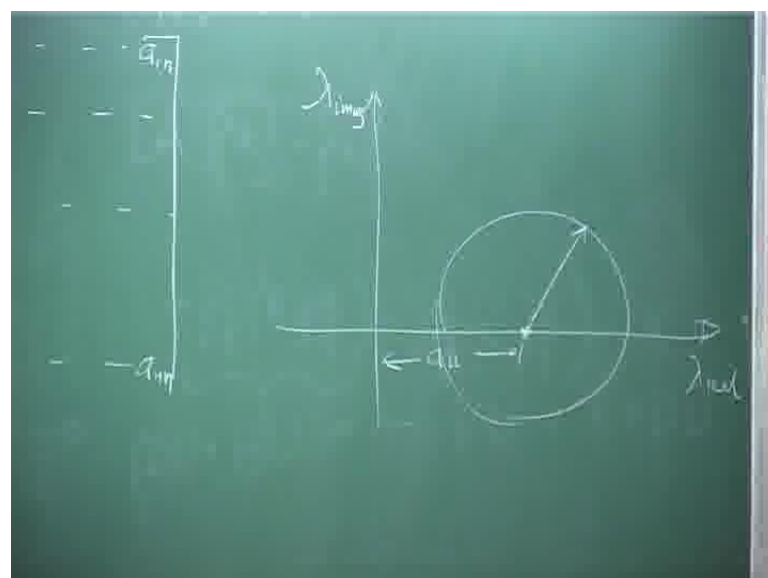
(Refer Slide Time: 48:10)

$$[A] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{bmatrix}$$

What we have is, let us say all are making it same. So, I can write down all these quantities  $a_{11}$ ,  $a_{12}$  and so on so forth till  $a_{1n}$  and  $a_{nn}$  (Refer Slide Time: 48:18). Let us have a full matrix like this.

The first theorem says that you can find all the Eigen values, - every Eigen values - of this matrix  $A$  would lie in the union of circular disc and this circular disc are formed in the way that the center is located at the diagonal, if I have  $a_{ii}$  so this is where the center would be located.

(Refer Slide Time: 48:58)





Basically, what we are talking about is, we are trying to find out the Eigen values of a. So, I will plot lambda imaginary and lambda real and what I would do? I will look at the ith entry of that and I will find out the center of that circle that is at a distance  $a_{ii}$  from the origin and then, the radius of the disc that is formed that is given by this. So, this radius is nothing but the sum of the magnitude of all the up diagonal term in that row.

(Refer Slide Time: 49:54)

## Gerschgorin Theorems

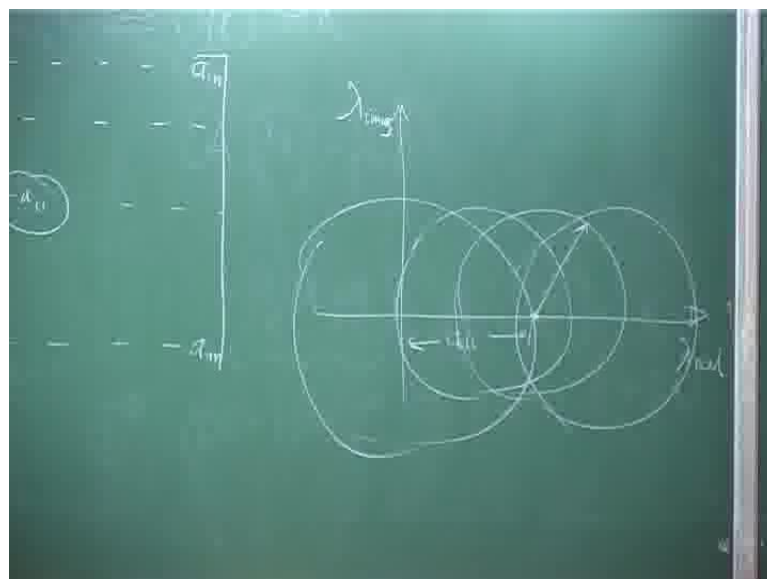
- Theorems due to Gerschgorin help in locating maximum and minimum eigenvalues of any matrix.

**Theorem 1:** Every eigenvalue of a matrix  $A$  (elements given by  $a_{ij}$ ) lies in atleast one of the circular discs with centers at  $a_{ii}$  and radii  $\sum_{j \neq i} |a_{ij}|$  ←

**Theorem 2:** If  $s$  of the circular discs of Theorem-1 form a connected domain which is isolated from the other discs, then there are precisely  $s$  eigenvalues of  $A$  within this connected domain.

Foundations of Scientific Computing, Chapter 10, p. 47/53

(Refer Slide Time: 50:08)



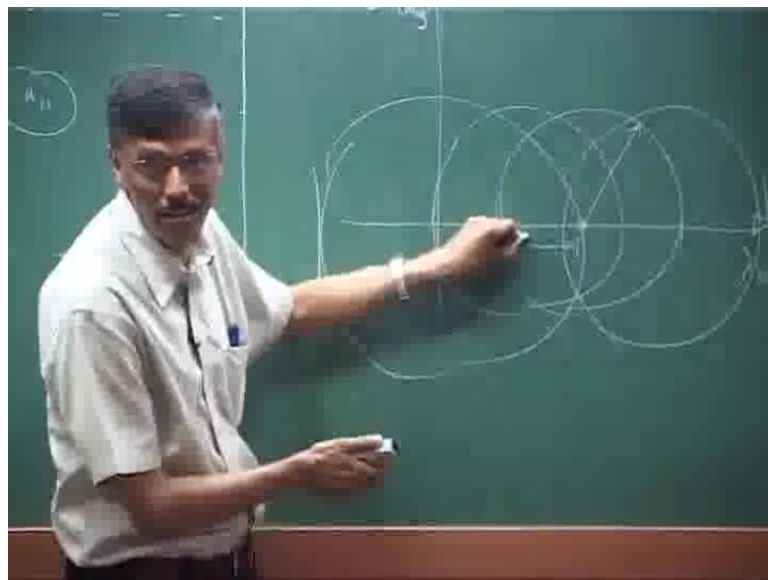
So, I could add  $a_{i1}$ ,  $a_{i2}$  and so on so forth and I sum it all up by taking its magnitude here and this excludes  $i$  equal to  $j$  point and then, I get the discs corresponding to this

row. What I do is, I keep plotting all possible discs, so I have this union of discs corresponding to each row.

Now, the maximum and minimum Eigen value would correspond to a global minimum and the global maximum. All we have to do is, take a look at that matrix and see what their entries are and perform this, construct this disc and this is not very difficult to do because looking at the diagonal I know where the center is. I can take the modulus of the half diagonal entries sum it up are the radius, so for that row I know the minimum is here maximum is here (Refer Slide Time: 50:57).

Then, I look at another row and compare this minimum and maximum. So, I keep on enlarging my union of the disc and find out what is the global, so that is a very easy task to do. In the present context, it is so much simple for you to do because it is tri-diagonal matrix. Your matrix has full of 0s, so all will have to worry about the diagonal term that is your center sum that two half diagonal terms that will give you the radius and what have we seen is some of this half diagonal term answers the diagonal term, you recall that we had written there what happens? We actually end up getting a disc which actually touches this point.

(Refer Slide Time: 52:01)



You will get the center in the middle diagonal term in here on the half diagonal term will fix it in the back. So that is rather very easy for want to do that. I think, we will continue

this discussion; I hope you will remember after your nice holidays. We are not done yet; we will talk about little more on ADI method. Now, you know that what ADI method is and we will talk about another method which is of contemporary interest in use that is called as the Multigrid method, we talk about that.