Numerical Solutions of Ordinary and Partial Differential Equations Prof. G. P. Raja Sekhar Department of Mathematics Indian Institute of Technology, Kharagpur

Lecture - 29 Matrix Stability Analysis of Finite Difference Schemes

Hello, today let us discuss matrix stability analysis of finite difference schemes. For example, if you approximate particular PDE by finite difference scheme, the same can be expressed in terms of matrix equation, and then when we introduce the corresponding errors, we may get corresponding error equation, which involve matrix and the amplification of a error in term should be dictated by the behavior of the matrix. So, let us visualize this for specific scheme, but to start with we may go a little bit of general notation, but then after sometime we will discuss specific methods.

(Refer Slide Time: 01:06)

Matrix Stability Analysis
Stability -> boundedness of the solution
Ut $\bar{u}_{i+1} = A \bar{u}_i + \bar{b}_i$ be an approximation LE CET let $u_{i+1} = n - i + o_1$ and $\frac{1}{n}$
idea: $\frac{1}{n}$ every to be bounded, one expects a

So, for example, when we say stability, this is related to boundedness of the solution. So, let your approximation can be written like this, then definitely the stability of the method depends on the matrix A, so what is the idea. The idea is for errors to be bounded, so when I say let so this is an approximation then the idea is for errors to be bounded, 1 expects a condition on A. Now, when we talk about condition on A, it is a matrix, where as the solution is a kind of factor.

So, there should be various kind of criterion, which is compatible. So, for example when we talk about suppose, he introduce error so definitely, the error is going to be magnified by a multiple of A at i th level. So, when we do that definitely, it depends on matrix A, but in what sense because it is matrix A, so how it is magnifying. So, this is related to the behavior of the matrix A with respect to various norms. So, 1 should do little bit of study on various types of matrix norms and vector norms and expect that you pick up a norm, which is compatible. That means, for the vector norm whatever we choose the corresponding norm we choose for the matrix.

(Refer Slide Time: 3:52)

So, some vector norms, so for me x means vector I am not putting a bar, so this simply 1 norm some time. So, for this case this is nothing but 5 then infinity norm max of the modules, so in this case this is 3 then square norm, max also taken i 1 to n. So, in this case it is route 13, so these are some vector norms.

(Refer Slide Time: 5:19)

Matrix Norms Martin Romes
 $||A||_2$ = watimum column sum of moduli of elements
 $||A||_2$ = watimum row Am of moduli of elements
 $||A||_2$ = $\sqrt{e(A^2A)}$, $A^H = (A^T)^T$
 $||A||_2 = \sqrt{e(A^2A)}$, $A^H = (A^T)^T$
 $||A||_2 = \sqrt{e(A^2A)}$, $A^H = (A^T)^T$
 $||A||$

Similarly, let us see some matrix norms, 1 norm maximum column sum of module of elements of A. Similarly, infinity norm maximum row sum of module of elements of A. So for example, A is given by 1 minus 1 minus 3 and 2, so if we choose such a matrix then so this will be maximum columns of sum of the module. So, this is going to be 4 and maximum row this is going to be 5 and also there is a other norm, so this is a where conjugate transpose and this is called spectral radius.

So, for example, let us choose A to B something like, so then what is a definition of spectral radius; we compute for example, I can give the Eigen values of this values of A. So, they are listed as therefore, is 11.86, so for a particular matrix we have to compute A h and then take the maximum Eigen value under route that will be the norm. So, this is general few introductions about atom norms and matrix norms.

(Refer Slide Time: 8:53)

Necessary and Sufficient condition fin stability
demain: $0\leq n\leq 1$, $0\leq k\leq T$
 $t_{i}=ih$, $i=0,1...N;nh=1$ $k_j = j k_j$ $j = 0 \cdots T$; $Jk = T$

Now, we talk about necessary and sufficient condition for stability. So, as I mentioned we may expect some condition on the matrix, so let our domain be this, this is a kind of domain. When we talk about this, we have to discretize x i is i h, then if we go for this accordingly we have N h equals to 1. So, generally 1 can go for a to b, but I have just here and 0 to 1. So, similarly t j is J k, J k should be our t, so when we have particular passive differential equation, then let us say in this case 1 space and 1 point and then our domain is something like rectangle like this.

We have to discretize as per this so we discretize. And similarly, we have to discretize so when we do this, we have to approximate by fund difference scheme of the given PDE and then talk about some condition of stability of the corresponding fund difference scheme. So, to start with let us consider two time levels that is method is involving j plus 1 and j.

(Refer Slide Time: 11:53)

D CET assume that the values at (j+1) and j are related by $\boxed{b_{i,1} u_{i,j,i+1} + b_i u_{i,j+1} + b_{i+1,j+1}}$ = $C_{i+1}u_{i+j} + C_i u_{i,j} + C_{i+1}u_{i+j}$
if the boundary values at $i = 0$ of N , j.10 and known = $C_{i+1}u_{i+j} + C_i u_{i,j} + C_{i+1}$ $num(x)$ at $i=1+n-1$ b_0 $\frac{a_{0,j+1}}{a_{0,j+1}} + b_1 u_{1,j+1} + b_2 u_{2,j+1}$ Heaven
= $c_0 u_{0,j} + c_1 u_{1,j} + c_2 u_{2,j}$

So, assume that the values at j plus 1 and j are related by f, so this is all j plus 1 to the left and this is equals to call this is as star. Now, if the boundary values at i 0 to n, j greater than 0 are known, so this is our star. Now, we had to talk stability of this approximation, so for example, for an explicit scheme b i minus 1 is 0 b i plus 1 is 0 and we have only u i j plus 1 to the left and remaining to the right something like that. So, if these boundary values are known at i 0 to n, corresponding to j equals to these are boundary values, so for all time level corresponding to these values, this would behave. So, for i and 0 and boundary, so then we have to run star at i equals to 1 to n minus 1 because achieving bounder values are given i 0 and n, so c i equals to 1.

Now, corresponding to a given initial boundary value problem, when we substitute i equals to 1, so this corresponds to boundary so this is known corresponding to the boundary condition. Now, this also known, so first equation these are to the left and these are to the right, so naturally if you supply the values at j times step, we can compute j plus 1.

(Refer Slide Time: 16:40)

 $\left[\begin{array}{c} 0 \text{ CET} \\ 0 \text{ T. KGP} \end{array}\right]$ $b_1 u_{i j+1} + b_2 u_{2 j+1} + b_3 u_{3, j+1}$ $i = 2$ = c_1 u_{ij} + c_2 u_{2ij} + c_3 u_{3jj} $i=N-1$ b_{N-3} $u_{N-3,i+1} + b_{N-2}$ $u_{N-3,i+1} + b_{N-1}$ $u_{N-1,i+1}$ = c_{N-3} $u_{N-3,j}$ + c_{N-2} $u_{N-4,j}$ + c_{N-1} $u_{N-1,j}$ $i = N$ $b_{N-2}u_{N-2,j+1}+b_{N-1}u_{N-1,j+1}+b_{N}$ = $c_{N-2}u_{N-1,j} + c_{N-1}u_{N-j,j} + c_{N}u_{N,j}$

Now, let us see for i equals to 2, what will happen? Now, if you continue like this, I would to write down further N minus 2, so this will be plus, this is equal to u, then N minus 1. So, this is slightly boring, but since I am following a more general notation as I mentioned the grid points 0 and n corresponds to boundary points, so these are corresponds to 1 to n minus 1. So the descries equation would be from 1 to n minus 1 of course, this is under assumption that we do not have derivative boundary condition.

Now in this, these are known due to the boundary values, now this can be put it in a form of a system so for example, since this is known this can be push to the right hand side and this stays right hand side. So, we can put it in terms of system starting from u 1 j plus 1 and u 2 j plus 1. So, let us see how we can do this so similarly, here we can push this to the right hand side.

(Refer Slide Time: 20:28)

So, if we do that we get b 1 b 2 because b 0 term is push to the right hand side, so then if we do this b n minus 3 n minus 2 N minus 1. So, these are 0 and trees, now this should be multiplied with which vector, so I would like to show it again. Consider the first one so b 1 is multiplying with u 1 j plus 1 u 2 is u 2 j plus 1 and this goes to the right hand side. So, since these two known, these two we would like to put it under one category and these are coming from $u 1 j u 2 j$.

First time when we run j equals to 0, these two values will be coming from initial conditions, so these are to be classified separately. So, if we do that b 1 must be multiplying with definitely u 1 j plus 1, so then b 2 must be multiplied with u 2 j plus 1. So, here we have a 0 because the next term u 3 j plus 1, so this gets multiplied and becomes 0, because in our first equation we have u 1 u 2. So, similarly, second equation u 1 u 2 u 3 so we can produce that this plus b 2 times, this plus b 3 times this. Therefore, the rest will be 0 so if we proceed like that similarly, the last this must be equal to i.

Hope we can manage here c 1 c 2 c 3 similarly, these are at j plus 1 level therefore, the corresponding vector here must be u 1 j u 2 j. We are left with additional terms for examples, the last case these are known they will be pushed to the right hand side and these are taken in the left matrix, these are taken in the right matrix. And one equation above all these are in the left matrix, all these are in the right matrix, so there would not be any term left out where, as in the last equation these two are separate and in the first equation these two are separate.

So, what we get here, we get another column at c $0 \text{ u } 0 \text{ j minus } b \text{ } 0 \text{ u } 0 \text{ j plus } 1$, I would like to write down c n u n j minus b n u n j plus 1 and 0 0 0 here. These are for the reason that I explained, so this is entire system and this system runs for i, so we have already run I, this is 1 to n minus 1 and then j will be 0 to some capital J. Now, we have to consider as a multistep process that means you plug in values at j compute and these are the known terms.

Now, as I mentioned the values at j plus 1 the values at j are multiplied by this, so if we can invert this and bring it to the right hand side then we can get the entire column threat j plus 1 level in terms of j th level. And hence the magnification depends on the resultant matrix, which will be that this matrix multiplied by inverse of this, which is what we have going to do. So, let us do that, this can be put it in the form.

(Refer Slide Time: 25:57)

$$
(M) \text{ can be put in a flow}
$$
\n
$$
B \overline{u}_{j+1} = C \overline{u}_{j} + d_{j}
$$
\n
$$
\Rightarrow \overline{u}_{j+1} = B^{1}C \overline{u}_{j} + B^{1}d_{j}
$$
\n
$$
\Rightarrow \overline{u}_{j+1} = A \overline{u}_{j} + f_{j} \qquad A = B^{1}C
$$
\n
$$
f_{j} = B \overline{d}_{j}
$$
\n
$$
\overline{u}_{j} = A \overline{u}_{j+1} + f_{j-1} = A (A \overline{u}_{j+1} + f_{j+2}) + f_{j-1}
$$
\n
$$
= A^{2} \overline{u}_{j+2} + A f_{j+2} + f_{j-1}
$$
\n
$$
= \overline{u}_{j+1}^{1} \overline{u}_{j} + \overline{u}_{j+1}^{2} + f_{j+1}^{2} + \overline{u}_{j+2} + f_{j-1}
$$
\n
$$
= \overline{u}_{j+1}^{1} \overline{u}_{j} + \overline{u}_{j+1}^{2} + \overline{u}_{j+1}^{2} + \overline{u}_{j+2} + f_{j-1}
$$

So, let us call this is some double star, double star can be put in a form b u j plus 1 second of notation equals u j. So, i is dropped and bar so this is just to denote as a vector d j. Each has a corresponding notation so what is a notation, this is b, this is c and this is d j and this is u j and this is u j plus 1. So, accordingly this implies a column vector now we have a nice matrix form, so u j intern. And further this can be expressed as so we started an iterative process, which is nothing but A square plus A plus. So, if you continue the iteration you are going to get f 0.

(Refer Slide Time: 30:26)

LE CET $\bar{u}_j = A^j \bar{u}_0 + A^{j-1} f_0 + A^{j-2} f_1 + \cdots + A f_{j-2} + f_{j-1}$

initial values

Let us perhab \bar{u}_0 to to \bar{u}_0 then
 $\bar{u}_0^* = A^j \bar{u}_0^* + A^{j-2} A + A^{j-2} f_1 + \cdots + A f_{j-2} + f_{j-1}$ - (0) define $\overline{e}_j = \overline{u}_j - \overline{u}_j^*$

(b) $40 \Rightarrow \overline{e}_j = A \overline{e}_{j+1} = A^2 \overline{e}_{j+2} = \cdots = A^j \overline{e}_{o_j}$ $j=1..T$ With respect to a *Mitothe norm* (*vector & motrice*)
With respect to a *Mitothe norm* (*vector & motrice*)

So, this one can absorb very easily just iterate and then we get this. So, we have u j equals, so this let us call some A and remember these are initial values because we are started iteration on j and when j is equals to 0, they are corresponds to initial values. Now, let us perturb u 0 to u 0 star because we are interested in studying the stability, we just perturb and see how the peturbation grows.

So, then if you perturb, perturbed also approximately satisfies the given scheme, so B then define the error vector e j as u j minus u j star. So, A and B give because from A and B, if it is subtract will get left hand side e j and right hand side each of this terms cancel and this is e_i 1, this will be a e_i minus 1. So, we get so this same whatever I mentioned, we get that difference will be a j, so we have a j e g e j is A power j e 0. So, as I mentioned, this is column vector where as this is a matrix, so we have to choose with respect to a suitable norm or compatible norm both vector and matrix norm less than or equals to A to the power j.

(Refer Slide Time: 34: 46)

D CET $||\overline{c}_{j}|| \le ||h^{j}|| ||\overline{c}_{s}||$

As pur Lax-Richtmyur, difference sented is stable when
 \exists a tre number M, independent of *j*, h and k \exists
 $||\overline{c}_{j}|| \le M$ $||\overline{c}_{0}||$ is amplification of initial
 $||\overline{c}_{j}|| \le M$ $||$

Now, we are almost close to the condition so we are writing now. As per Lax-Richtmyer difference scheme is stable, when there exist a positive number M, independent of j, h and k, such that so that means the positive- number m independent of j h and k, such that this happens. So, that means the amplification of the initially limited, so you can notice we are almost close to the condition.

Now what one would expect is we are almost getting condition on the matrix A, so let us see what could be the condition. So therefore, look for j, look for this condition or since A j equals to norm, A j minus 1 method is stable, if norm of A is less than are equals to 1. So, this is a condition for stability of course with a suitable compatible norm, so it is very fantastic you concerned to give finite difference scheme then put all j plus 1 to the left and all j th level to the right.

And you compute the co efficient multiplying j plus 1 and that matrix has been inverted and taken to the right. So, once it is taken to the right, you get a amplification matrix and for if you introduced a error, the corresponding amplification matrix bounded conditions, initial condition column vector is nullified, so you get a condition on the amplification matrix. And what is the condition it turned out to be the norm of the matrix must be less than or equals to 1. So, let us see observe this for a few exquisite cases, where we have various examples.

(Refer Slide Time: 39:03)

C CET As Consider the stability of
 $u_{i,j+1} = \lambda u_{i+1,j} + (1-2\lambda)u_{i,j} + \lambda u_{i+1,j}$, $i=1 \cdot N-1$ A = $B^{\prime}C = \begin{pmatrix} 1-2\lambda & \lambda & 0 & \cdots \\ \lambda & 1-2\lambda & \lambda & \cdots \end{pmatrix}$

So, let us do that example, consider the stability in this case. When we compare with our matrix you can see here, when we run i equals to 1 to n minus 1, the coefficient when i is 1 u 1 is getting survived, when i is u 2 is getting survived. So, compared to more general case, when i is 1 we have B 1 is 1, B 2 0 b 3 all are 0. Similarly, when i is 2 u 2 is there so B 1 is 0 B 2 is 1 B 3 so it turned to be that B is simply the identity matrix. So, comparing the mode general case B turned out to be, so let us see closely one more time so when we have B 1 is 1 B 2 0 B 3 all are 0.

Similar, when i is 2 u 2 is there so B 1 is $0 B 2$ is 1 B 3 is 3, so it turned to be that B is simply is the identity matrix. So, comparing the mode in general case B turned out to be so let us see closely one more time, so when i is 1 we have u 1 and we have u 0 u 1 j u 2 j. So, this is at jth level so will be pushed to matrix c and u 0 j will be pushed to column after B where, boundaries valves are known and we have u 1 j, so this is 1, so this is one level above.

We have this is 11 and 1 so we have this matrix then C if you these two will go the first equation. So, first equation the coefficient of u 1 j is 1 minus 2 lambda u 2 j is j lambda. In the second equation when i is 2 u 1 is lambda, u 2 is this, so c has a specified structure, we have structure like this. And A is nothing but b inverse c, so this will be the same because this is identity matrix. Now, in this case boundary valves are known u 0 j when i is $1 \text{ u } 0$ j i is n minus $1 \text{ u and } j$.

(Refer Slide Time: 43:47)

OCET bandary values us; and uni, are known for $j=1/2 \cdots$ when $|-2\lambda \ge 0$ i.e $0 < \lambda \le 1/2$ $\|\mathbf{A}\|_{\mathbf{A}} = \lambda + 1 - 2\lambda + \lambda = 1$ When $1-2\lambda 60$, $\lambda > 1/2$, $11-2\lambda 1 = 2\lambda - 1$ $||A||_{\infty} = \lambda + 2\lambda - 1 + \lambda = 4\lambda - 1 > 1$ This is the limit of the the same of the

So, another remark are known for now, if we observe closely we need condition on the matrix A so we have 1 minus sitting along the diagonals. So, when 1 minus 2 lambda is greater than and equals to 0, that is in these case these norms is lambda plus 1 minus 2 lambda plus lambda, which is 1 when 1 minus 2 lambda is less than 0, that is lambda greater than half we have.

So, in this case we have this, so in this case this is 4 lambda minus 1, 4 lambda minus 1 is greater than 1 therefore, the given method is stable for this range. So, the given method is stable for this range when lambda is greater than half, we failed to satisfy the condition therefore, it is not stable. So, the given method is nothing but an explicit method, which is Larsen method so it is stable in this range.

So, it is really greater to see how a particular finite difference method behaves and how the amplification matrix influence the growth of the error, so that if the norm of matrix A is less than or equals to 1, then the method is stable. So, alternatively there is another approach, of course still based on the Eigen valves and matrix analysis, let us see this alternative approach.

(Refer Slide Time: 47:00)

LE CET Alternate approach
 $\parallel A \parallel_2 = \ell(A) = \frac{m a \cdot |I|}{\sqrt{2}}$, $\frac{m}{\sqrt{2}}$, it eigenvalue of A A = $\begin{pmatrix} 1-2\lambda & \lambda & 0 & \cdots \\ \lambda & 1-2\lambda & \lambda & \cdots & \cdots \\ \lambda & \lambda & \lambda & \lambda & \lambda \\ & \lambda & \lambda & \lambda & \lambda \\ & & 0 & \frac{1-2\lambda}{\lambda} & 0 \\ & & & \lambda & \lambda \\ & & & & \lambda \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ & & & \ddots & \vdots \\ 0 & & & & 0 \\ 0 & & & & 0 \\ & & & & 0 \end{pmatrix}$

So, alternative approach two norm is defined has max of j u j where, u j is j th Eigen valve of A. So, in the present case A can be written as which is 1 minus 2 lambda. So, this can be written as plus I am going to split this minus 2, 1 then 1 minus 2 1, 1 minus 21 this must be lambda 1 minus 2 lambda apologizes. So, this must be 1 minus 2 multiplied by lambda, so all that we have done we have split this matrix into 2.

(Refer Slide Time: 40:36)

E CET $A = I_{N-1} + \lambda T_{N-1}$ TN-1 is such that the eigenvalues are $\alpha_g = -4 \sin^2 \frac{\lambda \pi}{2N}$

R = 1. N-1

hence the sigenvalues of A are $\mu_g = 1 - 4 \lambda \sin^2 \frac{\lambda \pi}{2N}$ $\|\mathbb{A}\|_2 = \max_{\mathbb{A}} \left| I - \mathbb{A} \wedge \widehat{\mathbf{b}}^{\mathbb{A}} \frac{\lambda \widehat{\mathbb{B}}}{2^{\mathbb{A}}}\right| \leq 1$ => -1 \leq 1 - 4 $\sin^2 \frac{\sqrt{m}}{24} \leq 1$, $8 = 1 \cdot N - 1$ $\lambda \in \frac{1}{2} \sin^2(\frac{N-1}{2N})$

Now, let us call this, where this matrix is i n minus 1 and this matrix n minus 1, then remark T n minus 1 is such that the Eigen valves are alpha j equals to minus 4 sin square j pie by 2 n. So, this j is not corresponding to time stepping, so you call some s, then s is just to make sure. Hence the Eigen valves of A are mu j equals to 1 minus 4 lambda mu s equals 1 minus 4 lambda sin square s pie by 2 n. Hence this max 1 minus 4 lambda sin square s pie, this must be equal to less than or equal to 1, that implies minus 1 less than 4 lambda sin square s pie by 2 n less than or equal to 1. From here we get lambda less than half so this you have to work out so this is just a algebra as h was to 0 and n was to infinite. As h was to 0 and n was to infinite, we have sin square n minus 1 pie by 2 n, this was to 1 therefore, hence lambda is less than or equal to half.

(Refer Slide Time: 52:34)

C CET have $\lambda \leq 42$ " the fid. scheme is contitiont
Lax. the = converges for och = 1/2

So, I am just giving the arguments therefore, the fid difference schemed is consistent and Lax theorem implies converges for. So, please work out some little bit of algebra because I just ran through, so we get this is the range where it is consistent and Lax theorem guarantees converges in the range. So, this is the matrix stability analysis where you can decide as stable stability of method and hence convergence using Lax theorem. So, there are also other methods I will discuss in coming lectures before that we may do couple of more problems.

Thank you.