

Foundation of Scientific Computing

Prof. T. K. Sengupta

Department of Aerospace Engineering

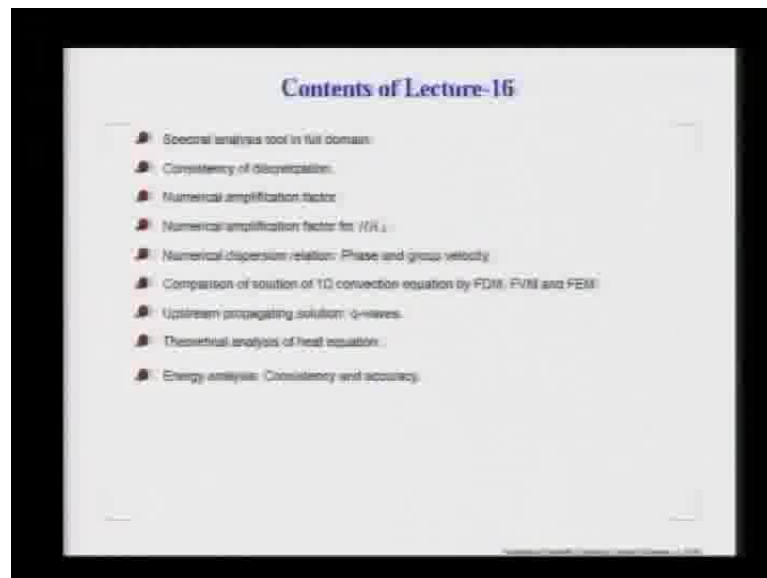
Indian Institute of Technology, Kanpur

Module No. # 01

Lecture No. # 16

Today's discussion on lecture 16 begins once again, by our discussion on Spectral Analysis tool. As we mentioned before, we would like to develop an analysis tool which is applicable for the full domain that would have different types of discretization for different points. So, in this context, we have already discussed in last lecture about equivalent wave number.

(Refer Slide Time: 00:16)



Once we have developed equivalent wave number, we are going to talk about consistency of discretization in terms of this equivalent wave number expression. Once we apply it to specific space time dependent equation, we can talk about numerical amplification factor.

We have seen already that Euler time discretization with various kinds of central schemes are unstable; that is why we would move over to numerical amplification factor for 4 stage Runge-Kutta method because we have already talked about the requirement of single step method as oppose to multiple step methods. Once we have talked about space and time discretization together, we are in a position to talk about numerical dispersion relation. Once we have the numerical dispersion relation, we will show how phase, phase speed, and group velocity could be computed from this numerical dispersion relation. We will do it specifically with the help of 1D convection equation and show the power of this analysis tool by comparing different finite difference, finite volume, and finite element methods, one by one.

We would like to bring one particular aspect of any discrete computing method. It is the existence of spurious upstream propagating solution and this is what has been called as Q-waves. So, this is something that we will be talking here, in great detail. This would basically conclude our discussion on discretization.

So, having finished our discussion on discretization, we will basically start our discussion on various solution methods. So, we will begin by solving parabolic partial differential equation. In this context, we will adopt the heat equation as an example of parabolic PDE.

We will begin our discussion of parabolic PDE solution method by theoretically analyzing the heat equation. Specifically, we would like to bring to your attention, the concept of physical instability versus numerical instability. That is why we need to have a former understanding of theoretical aspect of the solution. In this context, we may like to introduce an equivalent energy. Having done that, we will show that for a physically stable system, we cannot afford to accept numerical instability. In this context, we are also going to talk about consistency and accuracy of solution methods for PDEs. Then we should continue with our discussion.

We are actually in the process of developing an analysis tool which is in the spectral plane; that is why we call this as a spectral analysis tool.

(Refer Slide Time: 04:21)

Spectral Analysis of Computing

- Consider the one-dimensional convection equation:

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0 \quad (1)$$
- This has analytic solution that is non-dissipative and non-dispersive.
- The general solution is written in numerical framework, in the spectral plane by:

$$u(x_m, t^n) = u_m^n = \int U(k, t) e^{ikx_m} dk \quad (2)$$
- With the initial condition given by,

$$u_m^0 = \int U_0(k) e^{ikx_m} dk \quad (3)$$
- Numerically, we can **at the most** resolve any quantity up to the **Nyquist Limit** (k_{max}):

$$u(x) = \frac{1}{2\pi} \int_{-k_{max}}^{k_{max}} U(k) e^{ikx} dk$$
 where, for a grid of uniform size (h): $k_{max} = \frac{\pi}{h} \quad (4)$

Fundamentals of Fourier: Computing Spectral Analysis - p. 7/15

What we have been able to do is also analyze such a scheme in the full domain; that means, unlike what people have done earlier, people have developed methods where we could just simply look at the scheme - what happens in the interior. But you can notice here that we can find out effectiveness of discretization in terms of this k equivalent in a node-wise manner. So, for each and every j , I could evaluate this, the moment I decide to freeze upon the method of discretization through the choice of this C matrix. On the blackboard, we developed this.

(Refer Slide Time: 05:22)

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0$$

$$\frac{\partial u}{\partial x} \Big|_{C2} = \frac{1}{2h} [-u_{j+2} + 8u_{j+1} - 8u_j + u_{j-1}]$$

$$|K|_{C2} = \frac{1}{12h} [-e^{2ikh} + 8e^{ikh} - 8e^{-ikh} + e^{-2ikh}]$$

$$K_{C2} = \frac{1}{6h} [-\sin 2kh + 8 \sin kh]$$

We showed that if we take a second order central differencing scheme that gives k equivalent as $\sin kh$ by h . Then, from there, actually we drew a portrait of this effectiveness and plotted it in the non-dimensional wave number, ranging between 0 and π . On this side (Refer Slide Time: 05:33) we plotted k equivalent by k and **what** we notice that kh going to 0. What we are getting here? k equivalent by k here would be $\sin kh$ by kh (Refer Slide Time: 05:52). So, that is your $\sin x$ by x and we know the familiar property of the function that it just simply decays to 0 like this (Refer Slide Time: 06:03) and this is your value 1. So, ideally, what you would like to have is that all scales are resolved exactly, but discrete method shows that it is scale dependent. Depending on the value of k , you have different effectiveness.

What about this point? This point has to be equal to 1, why because this is the limit for which you are going from the discrete to continuum; h going to 0. If h goes to 0, then we reach this point (Refer Slide Time: 06:42). Then, by equivalent resolution, it should be exactly equal to the theoretical estimate. So, that is something that we have talked about.

(Refer Slide Time: 06:59)

Spectral Analysis of Computing (cont.)

- One can represent first derivative evaluated by any discrete method:

$$\frac{\partial u}{\partial x} = \frac{1}{h} [C] \{u\} \quad (5)$$
- In the spectral plane this can be written as,

$$\frac{\partial u}{\partial x}|_{x_j} = \frac{1}{2\pi} \int i k_{eq} U(k) e^{ikx_j} dk \quad (6)$$
- For spectral method: $k_{eq} = k$
- For discrete computing methods:

$$[i k_{eq}]_j = \frac{1}{h} \sum_{l=1}^N C_{jl} e^{ik(x_l - x_j)} \quad (7)$$
- For second order central differencing:

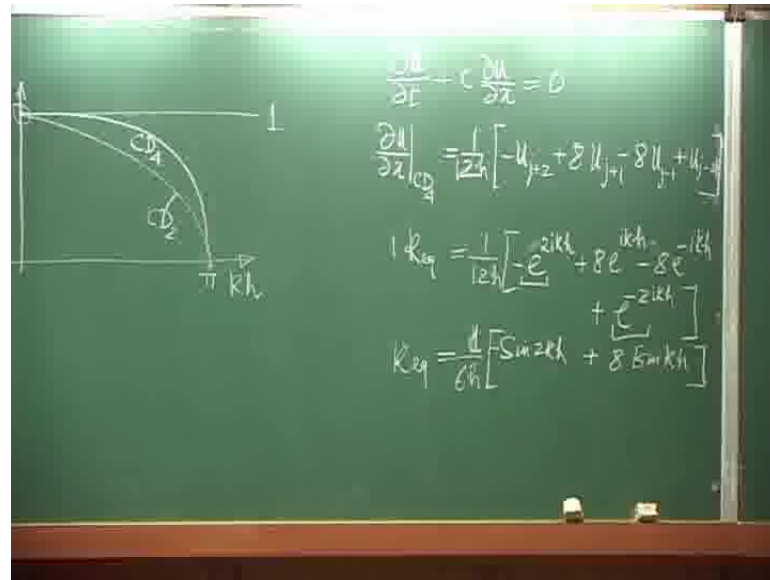
$$k_{eq}^{(2)} = \left[\frac{\sin(kh)}{h} \right] \quad (8)$$
- For fourth order central differencing:

$$k_{eq}^{(4)} = \left[\frac{\sin(kh)}{h} \right] \left[\frac{4 - \cos(kh)}{3} \right]$$

Foundations of Scientific Computing: Spectral Analysis - p. 9/15

Now, suppose you take care of this discretization of the first derivative in that convection equation, basically investigating this simple equation, we have seen what the second ordered discretization does.

(Refer Slide Time: 05:22)



So, suppose you do it by fourth order central differencing scheme, that I think you would recall, this was the expression that we had written - u_{j+2} plus $8u_{j+1}$ minus $8u_{j-1}$ plus u_{j-2} . So, once again we can calculate its k equivalent. You can very clearly tell me what this is going to be; k equivalent would be $1/12h$ and from here I will get e to the power $2ikh$; so, I will get minus e to the power $2ikh$. From here I will get $8e$ to the power ikh ; from here, I will have e to the power minus ikh ; the last one we will continue - e to the power minus $2ikh$.

So, you can see, things do happen pair-wise. So, you have here (Refer Slide Time: 08:45 to 09:36) appear with the opposite sign. So, you could club them together; each one of them will contribute 2. So, we could take this quantity out. So, this will be nothing but, I could also take i out. So, I will get here $\sin 2kh$ from this and that and from there we will get $8 \sin kh$. So, this is the expression for ik equivalent. So, you can write k equivalent by omitting that i . So, you will get that and do a little bit of simplification and you will get this expression.

So, **what** you find that the fourth order differencing is equivalent to the second order differencing quantity multiplied by this factor. This factor has a role to scale it up. What I mean by that is - if this is the figure that I have got for CD_2 , for CD_4 I will get something like this (Refer Slide Time: 10:05). So, it says that I have a much larger range of kh , over which my representation is more accurate.

So, this is the story with all explicit methods. As you keep on increasing its order, you keep seeing that this gives you better and better approximation. However, for all the cases, you would notice that this will go to 0 at the Nyquist limit.

[Noise] (Refer Slide Time: 10:36)

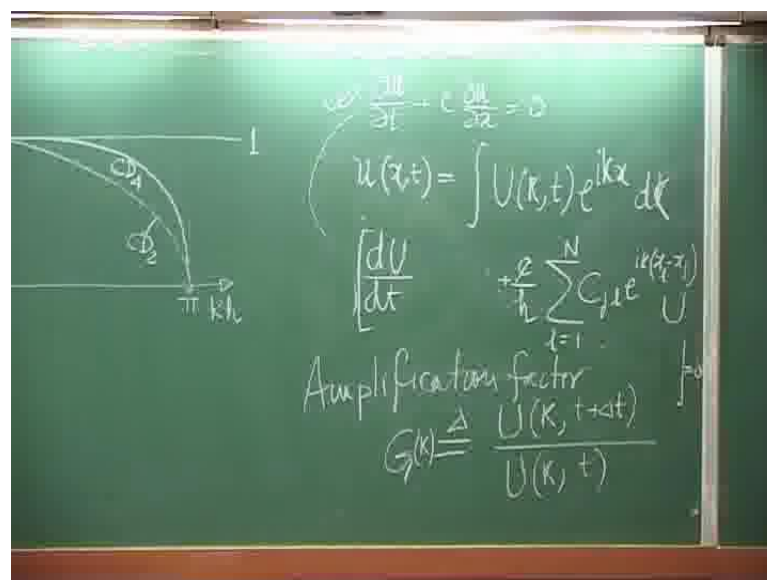
Yes. [Noise]

Pardon. Here? plus and that is why this has given us this (Refer Slide Time: 10:45) maybe I should have a minus sign here; so, I should have a minus sign there too. (Refer Slide Time: 10:57)

So, this is the story when it comes to discretization. So, what we are looking here is - what happens when we just discretize the spatial derivative term alone. Now, the story does not end there because what you end up doing is solving a particular equation where both space and time dependence come into play.

So, if I look at this equation, what we are going to do is again represent the unknown in terms of Fourier Laplace transform.

(Refer Slide Time: 11:54)



So, I will write like what I did yesterday; U of k and t into e to the power i k x d k. So, that is the way we are going to look at it. So, as you can see - this term (Refer Slide

Time: 12:20 to 13:20) will give you an integral dU by dt and this term remains as it is. This term will give us c by ... this is a lower case c . So, I will write it like this. What did we write this as? **In terms of** if I am doing it for the j th point, I will write it C_{jl} to the power $i k x_l$ minus x_j and l goes from 1 to n and this multiplied by u and the face part remains as it is. So, basically this is what we are getting from here to here via this spectral representation. Then, of course, it is true for the integrated quantity. So, the integrant itself must be equal to 0 and that is what you have it here.

The top equation is essentially what we have done - remove the phase part; remove the integral part; I have here is this (Refer Slide Time: 13:57). So, I have dU by dt plus c by h and the C matrix operating on the projection operator times U and so a little bit of a manipulation will you get you in this figure (())).

(Refer Slide Time: 14:17)

Spectral Analysis of Computing (cont.)

Using (6) in (1), one obtains,

$$\frac{dU}{dt} = - \left[\frac{c}{h} \right] \sum_{l=1}^N C_{jl} e^{ik(x_l - x_j)}$$

Define the **CFL** number as $N_c = \frac{c \Delta t}{h}$. If we perform Euler time integration, then the amplification factor $G(k) = \frac{U(k, t + \Delta t)}{U(k, t)}$ is given by,

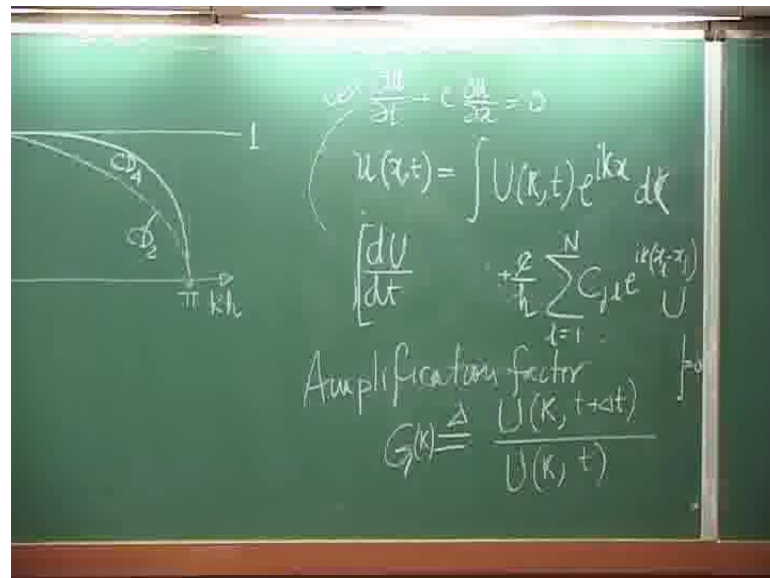
$$[G_j]_{Euler} = 1 - N_c \sum_{l=1}^N C_{jl} e^{ik(x_l - x_j)}$$

Foundations of Fourier Computing: Spectral Analysis - p. 9/11

Now, in most of your computational activities wherever convection is involved, you will always notice appearance of this parameter which is shown here in that square bracket; this is a non-dimensional quantity. This is what is called as the Courant-Friedrichs-Lewy number or CFL number. So, just simply remember it as a CFL number. What it basically tells you is, a kind of a non-dimensional quantity N_c which we will write and we will see that this is a fundamental independent variable that determines the property of the method.

Now, having defined the variable in terms of its Fourier Laplace transform like this, we can define what we will call it as an amplification factor.

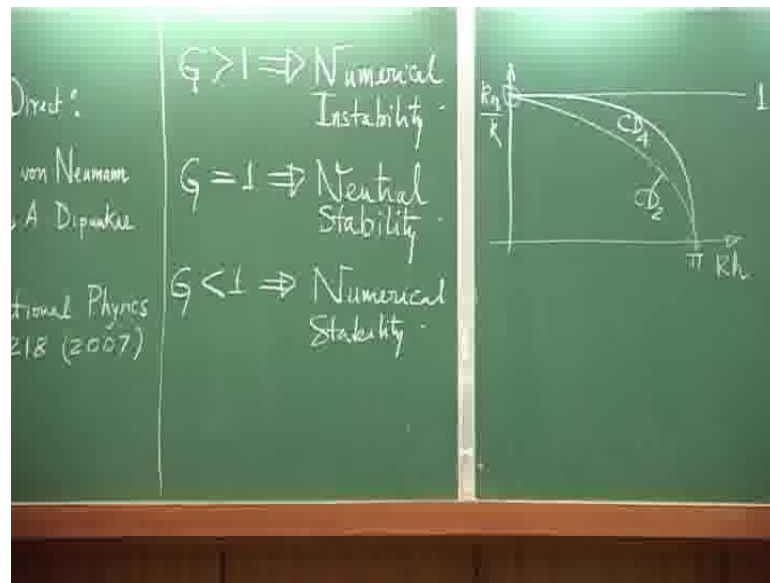
(Refer Slide Time: 15:22)



We will define it as U for that particular k , where we are looking at the solution at the advance time step divided by the solution in a k space at the old time step. So, in a sense, this is going to be a function of k .

You can very clearly see that in the limit of continuum when we take Δx equal to 0 Δt equal to 0, this G should be equal to 1; see easily, if Δt goes to 0, this limit goes to 1. What happens is a different story; we do a finite time step calculations and the moment we do that we deviate from its ideal value of 1.

(Refer Slide Time: 16:51)



So, in computation, irrespective of equation you will always expect G should be as close to 1 as possible. What does this mean? Look, if I have G greater than 1, what does it imply? It implies the solution is growing with time. So, I will call that as instability. Since this is an action of a numerical activity, I will call it as numerical instability. If G is equal to 1, then it is neither growing nor decaying; I will call that as neutral stability. When G is less than 1, we call this as numerical stability. So, with time, the Fourier Laplace amplitude will keep decaying.

There seems to be a lot of misconception among the practitioners of computing in the CFD community. I have noticed, time and again people tend to always think that you must have a stable algorithm; nothing can be far from truth. As you can see from the definition here, G should be equal to 1 and it does not matter what equation you are looking at. When I come to discussing parabolic partial differential equations, I will specifically pose a physical problem and I will talk about its physical instability and then relate that with the numeric.

However, irrespective of any equation that you are looking at, this is what we want. We should always aim for neutral stability; that is our ideal limit. Please do download this paper; this will have all those discussions given a little more in detail (Refer Slide Time: 18:56).

So, what happens is - once I have written it down like this, now suppose I perform a Euler time integration, so far we have been silent about what we are doing with the time integration; let us say I am performing a Euler time integration on this term.

(Refer Slide Time: 19:27)

$$\frac{dU}{dt} = \frac{U(k, t+\Delta t) - U(k, t)}{\Delta t}$$

$$u(x, t) = \int U(k, t) e^{ikx} dk$$

$$G_{\text{Euler}} = 1 - N_c \sum_j C_{jl} \left\{ \cos k(x_j - x_l) + i \sin k(x_j - x_l) \right\}$$

$$G_{\text{exact}} = 1 - N_c \sum_j C_{jl} \cos k(x_j - x_l)$$

$$G_{\text{implicit}} = -N_c \sum_j C_{jl} \sin k(x_j - x_l)$$

$$G(k) = \frac{U(k, t+\Delta t)}{U(k, t)}$$

So, what I would do? I would write it as u of k t plus Δt minus u of k and t divided by Δt .

Then, you can see, this is the outcome because there is a U sitting out there. So, I could pull it out and I get dU by U equal to minus of this N_c times this summation over this factor (Refer Slide Time: 19:49 to 20:18). What happens as a consequence, if I divide by U so this divided by U of k and t will give me G ; so G minus 1 will be equal to this factor; so G will be equal to 1 minus N_c into this factor.

What does it tell us? I have been telling you for a long time that this is a potentially a bad method to do Euler time integration. Why? You see, the C matrix is going to be of real entries. The way we discretize, you have noted various methods; some of them you have seen. C is a real matrix, but this phase function is complex. So, what we can do is I can take a modulus of this G and immediately you will notice that this is greater than 1. So, this modulus of G is greater than 1. What does it mean? That is an unconditionally unstable method.

So, you look worried. Tell me, you have any confusion? Let us work it out. So, I have $1 - N \Delta t$ and this C_j and this will be $\cos(k x_j \Delta t) + i \sin(k x_j \Delta t)$.

You can see this. So, you can see the real part $1 - N \Delta t$ (Refer Slide Time: 22:17) and you have the imaginary part which will be nothing but $i \sin(k x_j \Delta t)$ (Refer Slide Time: 22:28). Now can you see what I said?

So, you are now convinced that this modulus will be greater than 1 and that makes Euler time integration very undesirable; it will lead to instability. So this is that. What are the other better methods that we have?

(Refer Slide Time: 23:04)

Amplification Factor for RK4 Method

For the 4-stage Runge Kutta (RK4) method applied to the equation:

$$\frac{\partial U}{\partial t} = L(u) \quad (9)$$

The four stages of the method are:

Stage - 1: $u^{(1)} = u^{(n)} + \frac{\Delta t}{2} L(u^{(n)})$

Stage - 2: $u^{(2)} = u^{(n)} + \frac{\Delta t}{2} L(u^{(1)})$

Stage - 3: $u^{(3)} = u^{(n)} + \Delta t L(u^{(2)})$

Stage - 4: $u^{(n+1)} = u^{(n)} + \frac{\Delta t}{6} [L(u^{(n)}) + 2L(u^{(1)}) + 2L(u^{(2)}) + L(u^{(3)})]$

Let me tell you for some of the time integration methods that we have investigated, we have developed ourselves, we find this is a prime candidate which gives excellent property and this is your 4 stage 2 time level Runge-Kutta method.

So, let me explain how this method works and how this method is better in terms of numerical amplification factor. Suppose I have a space time dependent equation. So, I do all kinds of spatial discretization; put all those terms on the right hand side and call it as a L operator; so, that determines all your spatial independence. Then, we have this kind of an evolution equation $\frac{\partial U}{\partial t} = L(u)$. Well, please forgive me, this should be a lower case u ; this is not that capital u ; it should be lower case u .

So, $\frac{\partial u}{\partial t}$ by $\frac{\partial}{\partial t}$ is equal to L of u . By now, all of you are familiar; we have already done it when we were looking at solution of ODEs. In the 4 stage Runge-Kutta method we performed these 4 stages, having started with the solution at the n th level. We find out an intermediate stage solution which we call as u superscript 1. Having obtained that use that to calculate this function L of u here, and then, from there we calculate the second stage function U_2 ; then we have the U_3 ; finally, we collate all these intermediate stages into the next step solution that is u_{n+1} . This is all there in your notes.

So, in fact, you can notice that one of the brackets has gone wrong, up in the stage 2; anyway, 2 and 3 there is something wrong. So, basically, let us see what happens when we incorporate our spectral description and try to get the value of G for this particular time integration method.

(Refer Slide Time: 25:19)

Amplification Factor for RK4 Method

- For the 1D convection equation: $L(u) = -c \frac{\partial u}{\partial x}$
- For the numerical process:

$$\frac{\partial u}{\partial x}|_{Num} = \int i k_{eq} U(k, t) e^{ikx} dk$$
- We have also noted: $ik_{eq} = \frac{1}{h} \sum C_{jl} P_{jl}$. Therefore,

$$\begin{aligned} \Delta t c \frac{\partial u}{\partial x}|_{x_j} &= \int \sum C_{jl} P_{jl} \frac{c \Delta t}{h} U(k, t) e^{ikx_j} dk \\ &= \int N_c \left[\sum_{l=1}^N C_{jl} \right] U(k, t) e^{ikx_j} dk \\ &= \int A_j U(k, t) e^{ikx_j} dk \end{aligned}$$
- Thus, for the first stage of RK4 method:

$$\begin{aligned} u^{(1)} &= u^{(0)} - \int \frac{A_j}{2} U(k, t^n) e^{ikx_j} dk \\ &= \int U(k, t^n) \left[1 - \frac{A_j}{2} \right] e^{ikx_j} dk \end{aligned} \quad (10)$$
- Therefore, we can define: $U^{(1)} = U(k, t^n) \left[1 - \frac{A_j}{2} \right]$ (11)

Foundations of Parallel Computing: Spectral Analysis - p. 9/11

So, coming back to your 1D convection equation here (Refer Slide Time: 25:34), I can put this on the left hand side. If I do this, this quantity is nothing but your L of u that is your L of u for this. Now, we have also said that numerical description of the derivative with respect to x ; we could write it like this. So, if I write Δt times, there is a c out there and times $\frac{\partial u}{\partial x}$, we are going to get this. C is there, Δt comes when I multiply. As you can see in the previous stage, every stage I need to multiply by Δt here. So, here you can see, there is a Δt by 2, Δt by 2, and so on and so forth.

So, Δt is part of the story. So, we get once again that factor $c \Delta t$ by h ; that is what we called as the CFL number or N_c . So, $c \Delta t$ by h , we keep it up front here as N_c . Then, this is what we have done. This P_j is nothing but this quantity e to the power $i k \times l$ minus $h j$; I have just simply economized on space by writing that. Then you have to have the Fourier Laplace amplitude U_k of t and integrate over all possible case that is what you get. So, again let us economize in expression and call this whole thing here N_c times this summation C_j P_j ; let me call that as A_j .

What does the subscript j imply? j implies that we are looking at the phenomena at the j th node. So, that is what we are doing. Having done that this is our first stage. u of 1 is obtained in terms of the starting A point u n times this part. That is what we have to do and that happens to be minus A_j by 2 and this quantity. So, u n itself is U_k t n e to the power $i k \times j$. So, the whole thing can be written like this. So, I could write u of 1 as some kind of Fourier amplitude, capital U of 1. That capital U of 1 is nothing but evaluated at t n times this 1 minus A_j by 2.

So, this is the way that I will describe the first intermediate solution either in terms of 10, or in terms of its Fourier amplitude by this expression level.

(Refer Slide Time: 28:23)

Amplification Factor for RK4 Method

For the second stage of RK4 method:

$$\begin{aligned} u^{(2)} &= u^{(n)} - \int \frac{A_j}{2} U^{(1)} e^{i k x_j} dk \\ &= \int U(k, t^n) \left[1 - \frac{A_j}{2} \left(1 - \frac{A_j}{2} \right) \right] e^{i k x_j} dk \end{aligned} \quad (12)$$

Therefore, we can define:

$$U^{(2)} = U(k, t^n) \left[1 - \frac{A_j}{2} + \frac{A_j^2}{4} \right] \quad (13)$$

For the third stage of RK4 method:

$$\begin{aligned} u^{(3)} &= u^{(n)} - \int A_j U^{(2)} e^{i k x_j} dk \\ &= \int U(k, t^n) \left[1 - A_j \left(1 - \frac{A_j}{2} + \frac{A_j^2}{4} \right) \right] e^{i k x_j} dk \end{aligned} \quad (14)$$

Therefore, we can define:

$$U^{(3)} = U(k, t^n) \left[1 - A_j + \frac{A_j^2}{2} - \frac{A_j^3}{4} \right] \quad (15)$$

Now, we go to the next step. The next step follows in a similar manner because what do we do there? We take $u_n - \Delta t$ by 2 into L of u evaluated at the previous intermediate stage U of 1.

So, that is what I could do. I could write it in terms of its Fourier amplitude, capital U of 1. Again, this gives me capital U - this quantity, but U of 1 we have already written down as U into $1 - A_j$ by 2, but there, this is up front factor A_j by 2; so, that comes in here. So, the whole thing works out like U of k comma $t_n - 1 - A_j$ by 2 into $1 - A_j$ by 2. So, basically this whole quantity minus this space path is the Fourier amplitude for the second intermediate solution.

We proceed and obtain the third quantity, the third intermediate stage that is $u_n - \Delta t$. See, in U_1 and U_2 , we have Δt by 2; U_3 we have Δt ; so that is why we have just A_j ; otherwise, previously you were getting A_j by 2. So, U_3 is u of $n - A_j$ into U of 2 into this. U of 2 is in front; you plug it in there; this is what you get. So, that explains to you what this U of 3 is. So, this is an expression that helps you explain everything in the k plane.

(Refer Slide Time: 30:13)

Amplification Factor for RK4 Method

- Therefore, one step of RK4 method yields the next step solution by:

$$u^{(n+1)} = u^{(n)} + \int \left[-\frac{A_j}{6} - \frac{A_j}{3} (1 - A_j)/2 - \frac{A_j}{3} (1 - A_j)/2 + A_j^2/4 - \frac{A_j}{6} (1 - A_j + A_j^2/2 - A_j^3/4) \right] U(k, t^n) e^{ikx_j} dk$$

$$= \int U(k, t^n) \left[1 - A_j + \frac{A_j^2}{2} - \frac{A_j^3}{6} + \frac{A_j^4}{24} \right] e^{ikx_j} dk \quad (16)$$
- Therefore, we finally obtain:

$$[G_j]_{RK4} = 1 - A_j + \frac{A_j^2}{2} - \frac{A_j^3}{6} + \frac{A_j^4}{24} \quad (17)$$
- The amplification factor is a function of kh and N_c only, at each node.
- Since A_j 's are complex, G_j 's are also complex. Thus, apart from amplifying, G_j 's also will provide phase shift after each time step. What are the consequences of this phase shift? This is investigated next.

Foundations of Scientific Computing: Spectral Analysis - p. 9/13

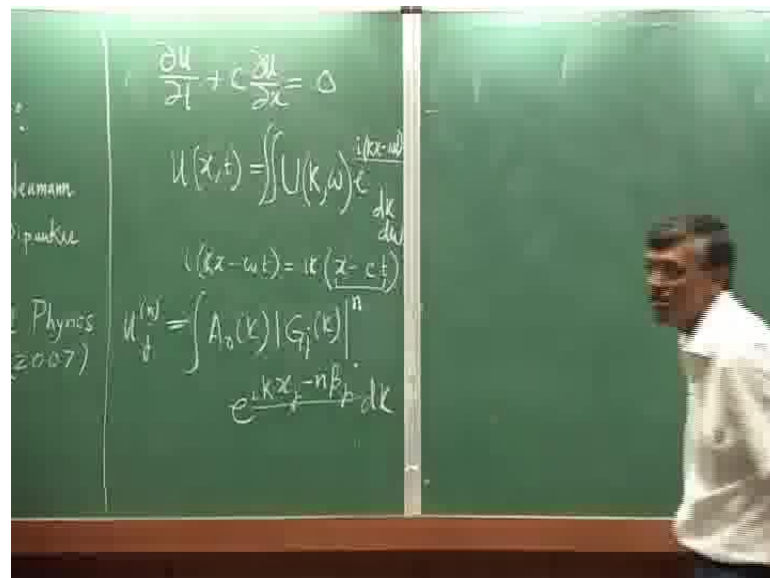
Now, having obtained all these quantities U_1 U_2 U_3 etcetera, you put it in the final collative stage where you get the solution at the new level in terms of the older value;

this is what you get; do a little bit of algebra and you get this. So, we get 1 minus A_j plus A_j square by 2 and so

So, what we get here for this RK4 method is this and as you can see, this A_j itself was a function of N_c and those P_j etcetera. So, that brings in the non-dimensional wave number, here k_h and N_c ; so, this G_j is going to be a function of k_h and N_c . Please do remember that A_j 's themselves are complex; so, G_j also will be complex. What it does? This sort of operation with G_j will not only amplify or attenuate, but it also will provide you with a phase shift. So, if I have a real quantity and an imaginary quantity, I can write it in terms of modular time step e to the power i phase. So, basically every operation or every time step would be equivalent to multiplying this previous time step solution with the amplitude plus a phase shift.

Now, you would be interested to know that what kind of a phase shift that you are getting? Because in solving this equation that we have started with, mainly the 1D convection equation, the solution is very straight forward.

(Refer Slide Time: 32:16)



The solution is, as if you recall, we would write it like this - if I write in terms of k and t , but let us say, now in terms of the frequency itself if I write, if I introduce frequency, then what will happen? Then, I will have e to the power $i k x$ minus ωt and $d k d \omega$.

So, that is what I will get and you can see the phase path. The phase path is $i k x$ minus ωt . So, if I take k out, I will get x minus $c t$. So, here actual solution shows the phase 2 change by this expression x minus $c t$. if I give you a solution at t equal to 0, at a subsequent time, you have the same solution, but it is shifted by $c t$ to the x minus $c t$. So, this is a kind of a phase shift.

Now, I want to know this G that we have uncovered here, for this RK4 method in 17.

What does it do? How is it related to the c that we are looking for in the exact solution?

(Refer Slide Time: 33:56)

Numerical Dispersion Relation

- One step of time integration by two-step method yields solution at the next step as:

$$u_j^{(1)} = \int A_0(k) G_j(k) e^{i k x} dk$$
- Since, $G_j(k) = |G_j| e^{-i \beta_j}$ where

$$\tan(\beta_j) = -[G_j]_{\text{imag}}/[G_j]_{\text{real}}$$

$$u_j^{(1)} = \int A_0(k) |G_j(k)| e^{i(kx - \beta_j)} dk$$
- Thus, every step of time integration shifts the phase of the solution at previous step by an amount: $-\beta_j$. One needs to ensure that this phase shift is according to the exact solution or not.
- After n such steps of time integration, we get the numerical solution as:

$$u_j^{(n)} = \int A_0(k) |G_j(k)|^n e^{i(kx - n\beta_j)} dk \quad (18)$$

Faculties of SCSE/CSE, Sardar Sarbajit College, Guwahati - 781 001

To understand this, I have demonstrated here, what happens. Suppose, I start from the initial solution which is given by the initial spectrum, $A_0(k)$ times $e^{i k x}$, then, I am looking for the solution at the first time step Δt ; that I will call, $u_j^{(1)}$; please do not confuse it with the first stage of RK4; this is what I am talking about time integration going from u_0 to $u_{\Delta t}$; so this is your $u_{\Delta t}$. That would be equivalent to multiplying by G of k ; that is the definition of our amplification factor.

Now, this G of k itself I have written here in terms of modulus and time, say phase shift. What is this phase shift? It is nothing but \tan^{-1} of $G_{\text{imaginary}}/G_{\text{real}}$ with a minus sign upfront.

So what happens is, I could then write this u, the solution at delta t would be the initial spectrum times the modulus of the amplification and e to the power i k x minus beta j.

Why do I write j? Because each and every point will have a different phase shift; that is what we have developed through this matrix theory that we can obtain this point wise. So, that is what we conclude here, that every step of time integration shift the phase of the solution obtained at the previous step, by this amount minus beta j.

One needs to ensure that this phase shift is according to the exact solution; we need to do that. So, suppose I perform n such steps, I have arrived at the solution at the nth time step; that would be again given by my initial spectrum.

There is a mistake. This G j of k, there should be a power n because every time I get 1, if I am doing n steps, there is n missing here. Please do understand that there is a mistake. Please do not quote me later that your note was wrong and that is why we have done it wrong.

So, what I am saying that you will have G j of K this has been operated n times. So, this exponent (Refer Slide Time: 36:27), this is not a superscript and then I have the phase e k. So, if I am doing this, every step I am getting beta j; so, I am going to get minus beta j this times d k. So, this n is missing there. So, please do note there correctly that this is incorrect variable.

(Refer Slide Time: 37:01)

Numerical Dispersion Relation (cont.)

- This actually provides us the numerical dispersion relation as, $n\beta_j = k c_N t$ (19)

where c_N is the numerical phase speed and the above is equivalent to the **actual numerical dispersion relation**:

$$\omega_N = k c_N \quad (20)$$

- This has to be contrasted with the wrong dispersion relation used so far in: Trefethen (1982), Sengupta et al. (2003), Lele & Colonius (2004) and De & Eswaran (2006).
- The correct dispersion relation as explained here, was presented in a series of papers: Sengupta & Dipankar (J.Sc. Comp., 2004), Dipankar & Sengupta (J. Comp. Phys., 2006); Sengupta *et al.* (J. Sci. Comput., 2006) and Sengupta *et al.* (J. Comp. Phys., 2007).

Then what happens? Now, having obtained this expression, you notice - compare this with this here (Refer Slide Time: 37:10). If I look at this phase relationship, with this phase relationship, I can see something emerging; this β_j is somewhat related to this ω here.

(Refer Slide Time: 37:24)

The chalkboard contains the following handwritten content:

- Left side:**
 - $\frac{\partial}{\partial x} = 0$
 - $\sum_{j=1}^N \frac{\beta_j}{\Delta t} = k C_N$
 - $C_N = \frac{\beta_j}{\Delta k \Delta t}$
 - $\frac{\partial N}{\partial t} = \frac{\beta_j}{(\Delta k \Delta t)} = \frac{\beta_j}{\omega \Delta t}$
- Right side (Q-Waves!):**
 - $\omega = k C \rightarrow E$
 - $\omega_N = k C_N \Rightarrow N$
 - independent variable: k
 - $\omega_N = k C_N$
 - $V_{gN} = \frac{d\omega_N}{dk}$

What is n ? n is nothing but the time step. So, that will be like your t_n by Δt . So, what happens then? n times β_j that is what we are seeing here,; that is going to give me something like this - $i k c t$.

So, I will write that as, there is an i here also and $n \beta_j$ should be equal to $k C t$, if everything was fine and nice. However, we have already seen that doing discrete computation means sacrificing something. What is that something? k is still our independent variable. So, we are keeping that as our point of reference. We are going to see that this C which was to be a constant in the exact solution numerically does not remain so; it does not remain so.

So, if I replace this n by what I have written there - t by Δt , then you can see this t will cancel and what we are finding here? An expression for C_N - that is nothing but β_j by $k \Delta t$. From the exact solution, we have noted ω equal to $k C$. So, this is our exact solution. But numerically what we are doing? k still is the independent

variable; C does not remain the same; this also, this is your numerical dispersion relation. See, we are relating ω with k . That is what we define.

At this stage it may appear it is a very simple thing. Tell you what that people have been doing things wrongly for decades? Including us, as you can see here, it started with professor Trefethen's work in 1982; then Lele Colonius also and this professor Eswaran from mechanical copied our work wrongly and they also ended up doing wrongly.

So, what you understand here is that your independent variable is k . What was the mistake people we are doing wrongly before? You know what people we are doing wrongly before; they were just simply writing ω as what we have done in the beginning of the class, change k to k equivalent, and then say multiply by C . So, this is wrong whereas, this is the correct way of expressing the dispersion relation (Refer Slide Time: 41:31).

What does it mean? I mean are we just simply nit picking? No. It is very profound because what we are seeing here is that this C is now a function of k , whereas in your exact solution, C was a constant. That is why we talked so glowingly about that say non dispersive, non dissipative solution. Here, what we are seeing is an act of discreet computing. We are getting a numerical phase speed C which is a function of wave number. So, that means what? Different k component will send their crest at different rates. So, this is really a big development which was first pointed out here. Well, we have tried to convince people with most of publications there, as you have noticed.

(Refer Slide Time: 42:40)

Numerical Dispersion Relation (cont.)

The consequence of $c_N \neq c$ is profound in terms of error and stability analysis, as will be discussed in details later. In fact, this proves the von Neumann Stability analysis wrong- one of the major results in computing considered so important that it was kept classified for many years in 1940s!

From (19), one can obtain:

$$\left[\frac{c_N}{c} \right]_j = \frac{\beta_j}{\omega \Delta t} \quad (21)$$

$$\left[\frac{V_{gN}}{c} \right]_j = \frac{1}{h N_c} \frac{d\beta_j}{dk} \quad (22)$$

Thus, we are in a position to obtain the numerical group velocity and numerical phase speed for the solution of one-dimensional convection equation. This is shown next.

Foundations of Scientific Computing: Spectral Analysis, p. 1115

Continuing about discussion that C_N is not equal to C is really profound in terms of error and stability analysis; there is a nice history about this. When during the Second World War, this group of people was in the Manhattan project in New Mexico developing the atom bomb. One of their main stalwart mathematicians helping them was Von Neumann.

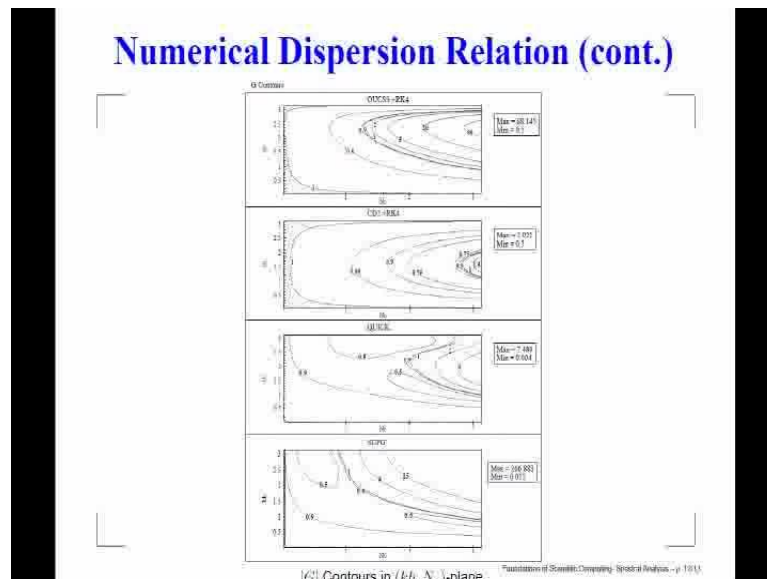
Von Neumann actually developed an error analysis or stability analysis. This was based on this assumption, a wrong one (Refer Slide Time: 43:30). At that time during the war, people are secretive about what goes on in research front; they classified the work; so, people did not know what was the work, but everybody knew that Von Neumann has done something which revolutionizes computing; explains a lot of features of computing. So, it was only I think in 1947 and 49, some papers started coming out, but as you can realize, that work was wrong and we actually first brought it out. We corrected Von Neumann's error; that is the main thing about this work.

What we notice out of all these exercise? I have got an expression for C_N . So, what I could do? I could define a non-dimensional quantity which I will call c_N by c . So, that would be β_j and I will have $k c$ into Δt . $k c$ is ω ; so, β_j by $\omega \Delta t$. So, that is what we have written here in 21. So, this is your consequence of numerical activity that you do not see c_N by c equal to 1, but it becomes β_j by $\omega \Delta t$. So, what happens?

So, you choose a method for spatial discretization. You choose a method for temporal discretization. You obtain the value of G which has a real path which has an imaginary path; you find out what is the imposed phase shift from this G , and that determines how far it is for one. Having given you this expression here (Refer Slide Time: 45:33 to 46:08), you can immediately calculate its numerical group velocity, which I will call as V_{gN} , which will be nothing but $d\omega_N/dk$. So, if I plug that in there, I will get this equal to c_N plus $k d c_N / dk$. If everything was nice and fine, you should have seen that these two should have been the same; V_{gN} equal to c that is an exact solution.

But the very fact that numerical calculation makes C_N non constant function of k , adds on this part that is the source of numerical dispersion. In fact, lot of calculation goes on in the literature, where people do claim that they have done this and that; they are essentially source of this Fourier's dispersion; that is inherent with all numerical methods; you cannot just simply wash them away until unless you choose your methods and parameter very carefully. That also tells you that at this point in time with the type of computing power that we have, we cannot solve the equation in a direct sense. We will always have to leave some kind of error.

(Refer Slide Time: 47:12)



What is the error? How this error is contributed? As we go along, we will explain more. At this point in time, I will show you some results.

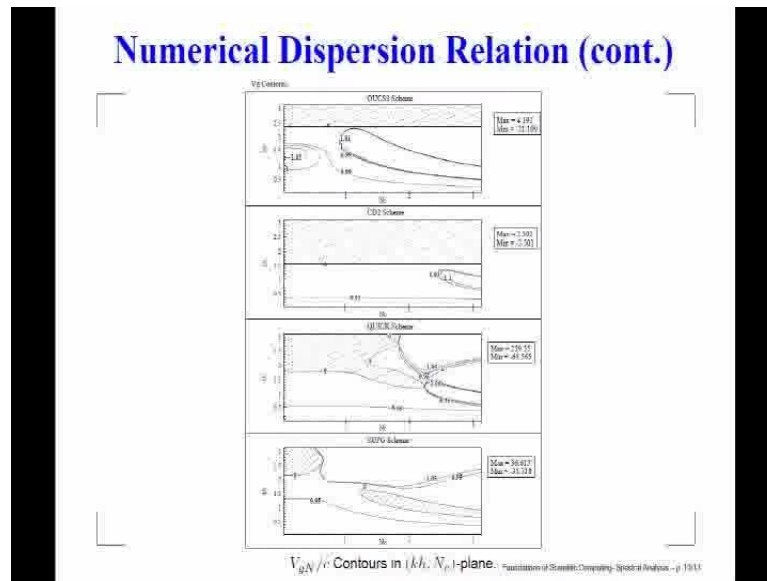
This figure may not make tremendous sense, except that in this figure what we have done? We have plotted these G contours; that modulus G that we talked about; the method in the second figure that we have is a CD2 method along with RK4. We have marked the region here with dash line where, your G is mod G is 1.

Recall, I wrote down that to do a correct calculation we must have neutral stability. So, This have been plotted in this plane; on the x axis I have N_c - the CFL number, on the y axis I have $k h$. It tells you that to do an error free calculation coming from numerical amplification consideration, you need to keep your Δt very small; that means N_c very small, so that you remain in the numerically neutrally stable region.

There are other methods. I will talk about these methods, but just simply know that this is the finite volume method, this is the finite element method, and unfortunately, neither of these two methods which are very popular very much in use; may be tens of, thousands of people use.

Then, as you can see, they do not have any G equal to 1 region. So, all they get is lot of... well you have solutions here; this is G equal to 1 line here; on this side, you have totally damp solution. So, as you keep integrating, your solution amplitude will come down; on this side, you have unstable path - this pocket. The same thing happens here with this finite element method called Petrov-Galerkin method; some of you may have taken a course; you know what it is; what you notice that G equal to 1 line is here. On this side, it is a completely damped solution, and on this side you have amplified solution. So, this is about the story of G .

(Refer Slide Time: 49:46)



You need to also know, what the numerical dispersion is. We have talked about this; b G n by c ; so, if we plot that again in N c at k h plane, we notice some very interesting feature. Let us keep our attention focus once again on this second figure because that is what we are quite familiar with now. We noticed that that V g N by c - this contour line, I suppose this line is 0.98. So, basically, even if you look at very small range of k h , here you are already started getting 2 percent dispersion; instead of 1, it is 0.98. This line here corresponds to 0.98, but interestingly enough, look at this line that is a very fascinating and interesting line that is V g N by c equal to 0.

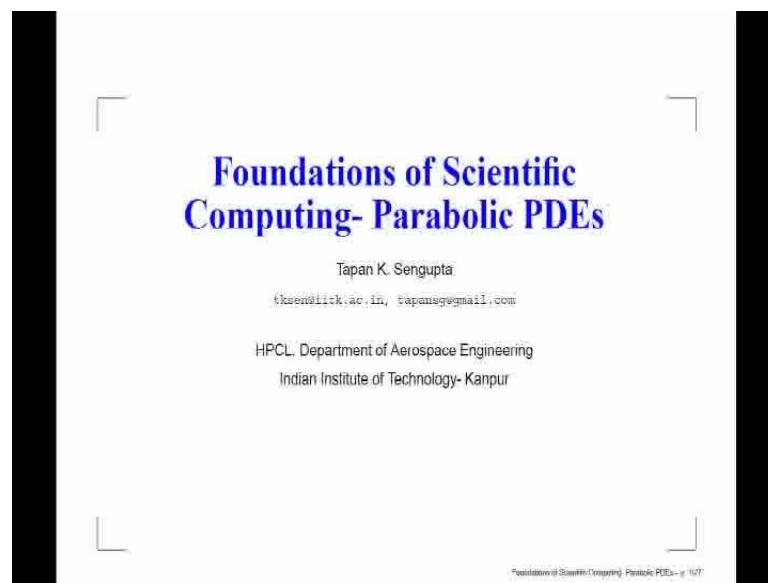
What does it mean? Below that line, you have V g N by c positive; above it is negative. The equation that we started solving, solution should have propagated from left; if c is positive it goes from left to right. If my numerical dispersion relation is such and I am in this part, solution will go in the wrong direction.

This feature is vaguely understood, but this is where we have actually quantified it and put them across for different types of methods currently in use. This is one of the methods that we have developed. You can see that this value is actually π by 2. So, for any k h value which is greater than π by 2, they will go in the wrong direction. Since these are not physical waves, somehow this nomenclature has stuck this type of spurious solutions are not called p -waves, but q -waves. So, q -waves means spurious upstream propagating solutions.

As you can see, almost every numerical method has this kind of a feature. So, you cannot just simply say that I have 1 method superior to other; in fact, the case for finite volume and finite element is really pathetic because they not only attenuate the solution, their dispersion relation property is also equally bad.

So, I think we would conclude here. I still have not told you about the error analysis path which will come little later.

(Refer Slide Time: 52:51)



You please download this paper and then I will come back shortly, but let me now go to the next topic that we would like to discuss.

(Refer Slide Time: 53:10)

Theoretical Analysis of Heat Equation

- Consider the one-dimensional heat equation:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \text{ in } 0 \leq x \leq 1 \text{ \& } t \geq 0 \quad (1)$$

- Solution of this requires an initial condition:

$$u(x, t=0) = f(x) \text{ for } 0 \leq x \leq 1 \quad (2a)$$

- Also, one would require boundary conditions, that could be given by the following Dirichlet condition:

$$u(0, t) = p(t) \text{ and } u(1, t) = q(t) \quad (2b)$$

- This is a **Parabolic PDE** with the characteristic:

$$t = \text{constant}$$

- One would like to investigate, what the theoretical solution does at $t \rightarrow \infty$?

That is basically going back to classical thing that any computing course tries to teach you - how to solve different types of PDE's. So, let us begin with parabolic PDE's because that is how historically it all began.

Well, once again with our formal practice, we start with some given equation. So, let us look at say 1 dimensional heat equation in a domain, finite domain x , varying between 0 and 1, and for all time we want to get the solution. It is a space time independent solution; so, you require an initial solution here; that is given by this function f of x ; in addition, it is a bounded domain problem between 0 and 1; so, you need to prescribe boundary conditions; the boundary conditions could be time dependent; so, that is why we have written them as p of t and q of t .

Now, we have already studied that. We know that the characteristic is t equal to constant; that is how the information propagates.

(Refer Slide Time: 54:32)

Asymptotic Energy Analysis

- For the solution $u(x, t)$, construct a non-negative energy functional:

$$E(t) = \int_0^1 u^2 / 2 \, dx \quad (3)$$

- The time-rate of this energy functional can be written as,

$$\begin{aligned} \frac{dE(t)}{dt} &= \int_0^1 u \frac{\partial u}{\partial t} dx = \int u u_{xx} dx \\ &= \int_0^1 \left[\frac{\partial (uu_x)}{\partial x} - u_x^2 \right] dx \\ &= q(t) u_x(1, t) - p(t) u_x(0, t) - \int_0^1 u_x^2 dx \end{aligned} \quad (4)$$

- Various sub-cases can be considered:

- If $p(t) = q(t) = 0$, then we do not have any input to the system through boundary condition and,

$$\frac{dE}{dt} = - \int_0^1 u_x^2 dx \leq 0 \quad (5)$$

Foundations of Scientific Computing, Pankaj PATEL - p. 307

Now, we would like to first discuss, what the property of the physical solution itself is for very large time because if I do not know that then I do not know what I am computing; so, first and foremost, I would like to know what the solution is doing. To understand that we define a quantity - a non-negative, functional, which I represent as energy. So, let us call this capital u of t as u square dx. So, this is a positive function.

We want to find out how this quantity changes with time. So, what you do is you have the definition, differentiate it with respect to time; then you will get $u \cdot \frac{\partial u}{\partial t} = u \cdot u_{xx}$; so I have got this; then I can do it a little jocularly here and that is what I am going to get $\frac{\partial}{\partial x} (u u_x) - u_x^2$. So, if I do that this is exact differential. So, I can integrate it out. With the help of those boundary conditions at x equal to 1 and x equal to 0, the first part gives me these two solutions; whereas, the last part, I keep it as it is; that is minus of u_x^2 dx.

Now, many a times, most of the times when you have thought this, any of these equations, especially the heat equation, you have most of the time told - let us look at something; we give some initial temperature distribution and then see what happens. So, there is nothing from the boundary. So, if I do that this p of t and q of t are 0; then this part is not there; then what happens? I have dE by dt is equal to minus of dx . So, it is a strictly negative quantity. So, what does it say? That if I do not do anything through the boundary, then the energy is going to decay with time.

So, that is a very nice feature of a physical solution. We do not like to consider a physical case where energy grows unbounded. There would be such problems of physical instability, but that is not what we have talked about. Here, we are talking about a benign case where, solutions do not block. What it shows? If I have a rod, I create some kind of a heat; heat distribution at t equal to 0.

(Refer Slide Time: 56:57)

Asymptotic Energy Analysis (cont.)

- The 'energy' of the system decays with time - a **physically stable system**. One should be able to compute it indefinitely.
- If $\frac{dE}{dt} > 0$, then we have a **physically unstable system**. This cannot be computed for long time.
- **Numerical Stability Requirement:** Computed 'energy' of a physically stable system must remain bounded.
- Other numerical requirements:
 - (i) Accuracy and
 - (ii) Consistency

Foundations of Scientific Computing, Parviz M. R. - p. 507

What happens subsequently? It says, where you look at the solution, the energy will continuously keep coming down. However, you can realize by a judicious choice of this function p and q and t , we can do lot of interesting things.

So, in the second part, I will show you some interesting thing that we have done very recently. To show you that even for this parabolic equation, heat equation, you can actually generate wave solutions; we will do that but later.

So, we have realized that when we do not have any boundary excitation, the energy of the system decays with time. So, it is a basically physically stable system; then one should be able to compute it indefinitely; there should not be any problem until unless your method is wrong.

If your energy increases with time, then we have physically unstable system; you cannot compute it. You will see that other physical processes will intervene and you will never get a situation where solution goes unbounded because some of the energy has to come.

Various processes like what we studied in case of solitons, you realize that there was this competition between focusing and dispersion - that got us a steady state solution.

So literally speaking, in physics, you will never come across a continuously unstable system; it will be unstable, but then it eventually saturates because of other processes. However, coming to our numerical stability requirement, we need the energy of the physical system to remain bounded; not only that we also need these two quantities which would make sense; we need the solution to be accurate and we need the method to be consistent.

See, we have defined something like this - the energy is going to decay with time. I will introduce you to a method which was introduced with a lot of fanfare and people actually used for nearly 30-40 years; even you can go to the search engine and find that there are still some people using it this method called Du Fort-Frankel method.

What happens? It does not follow the principle of the physical solution. Such methods are called inconsistent methods. So, we must be concerned about consistency. So, I think it is a nice place to stop here.