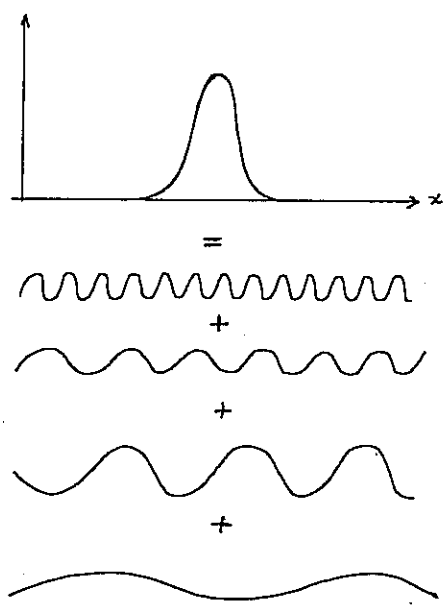
Lecture No. 6

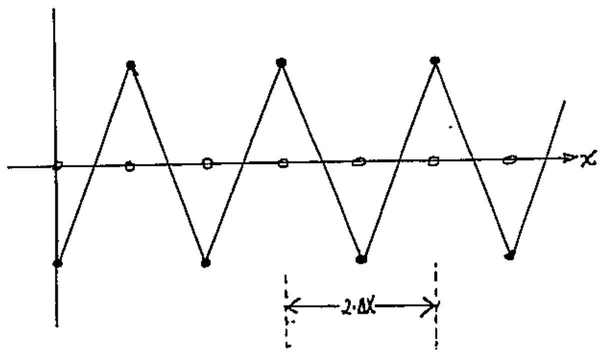
Lapidus and Pinder, 108-185; Smith, 43-49; 67-75.

General Considerations for Stability and Error Analysis

- The discrete solution at any point in time includes errors due to the initial numerical approximation to the specified i.c. as well as truncation errors and roundoff errors.
  - Truncation errors originate from the discrete representation of each differential term in the p.d.e. The difference between the continuum form of the p.d.e. and the discrete form of the p.d.e. represents the truncation error.
  - Roundoff errors arise from the computers inability to provide an infinite number of decimal places of precision to represent a number.
  - Truncation errors always build up in time (i.e. as the solution time marches forward from time level t_j to t_{j+1}).
  - Roundoff errors may or may not build up in time.
- The discrete solution can always be represented by a Fourier series.



- The shortest resolvable wavelength of any numerical solution is $2 \cdot \Delta x$.



- Insufficient numerical resolution (i.e. not enough nodes in space and/or time) can result in discrete solution components existing at the "short" wavelength ($2 \cdot \Delta x$) portion of the Fourier spectrum.
- Roundoff error always results in discrete solution components existing at the short wavelength portion of the spectrum.

<u>Stability</u>

Stability relates to the stable decay or unstable growth of component of the discrete solution (which includes errors in representing the specified i.c. as well as truncation and roundoff errors). Whether the various (Fourier) wavelength components of the discrete solution are amplified or damped depends on the nature of the difference equations not the p.d.e.!

- For unconditionally stable schemes, all wavelength components of the solution are always damped. Typically implicit schemes are unconditionally stable (i.e. there are no restrictions required to make the scheme stable).
- For unconditionally unstable schemes, some or all wavelength components of the solution are always amplified. This makes the numerical solution useless.

- For conditionally stable schemes, the numerical scheme will be stable under a required set of conditions.
- Stability is a necessary but not sufficient condition for a numerical scheme to be accurate.
- An unstable scheme is not convergent.

Heuristic Approach to Stability

- Illustrates how a discrete solution or a Fourier component of a discrete solution behaves in the time marching process.
- The approach does not provide information regarding the bounds on stability unless every possible combination/situation is tried out!
- Thus we will examine what the difference equations do to some discrete solution.

Application of the Heuristic Approach to Stability

Let's examine the growth/decay of a high wavenumber (i.e. short wavelength or near $2 \cdot \Delta x$) initial condition (i.c.) represented by one non-zero value at a point in space.

We will consider the growth/decay of this i.c. for the explicit discretization to the time dependent diffusion equation:

Recall that the explicit FD discretization for the time dependent diffusion equation is:

$$u_{i,j+1} = \left[\frac{\Delta t D}{(\Delta x)^2}\right]u_{i+1,j} + \left[1 - \frac{2\Delta t D}{(\Delta x)^2}\right]u_{i,j} + \left[\frac{\Delta t D}{(\Delta x)^2}\right]u_{i-1,j}$$

Letting:

$$\rho = \frac{\Delta t D}{(\Delta x)^2}$$

Leads to:

$$u_{i,j+1} = \rho u_{i+1,j} + [1 - 2\rho]\, u_{i,j} + \rho u_{i-1,j}$$

<u>Case 1</u>

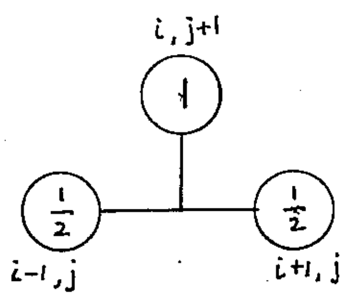
Letting;

$$\rho = \frac{1}{2}$$

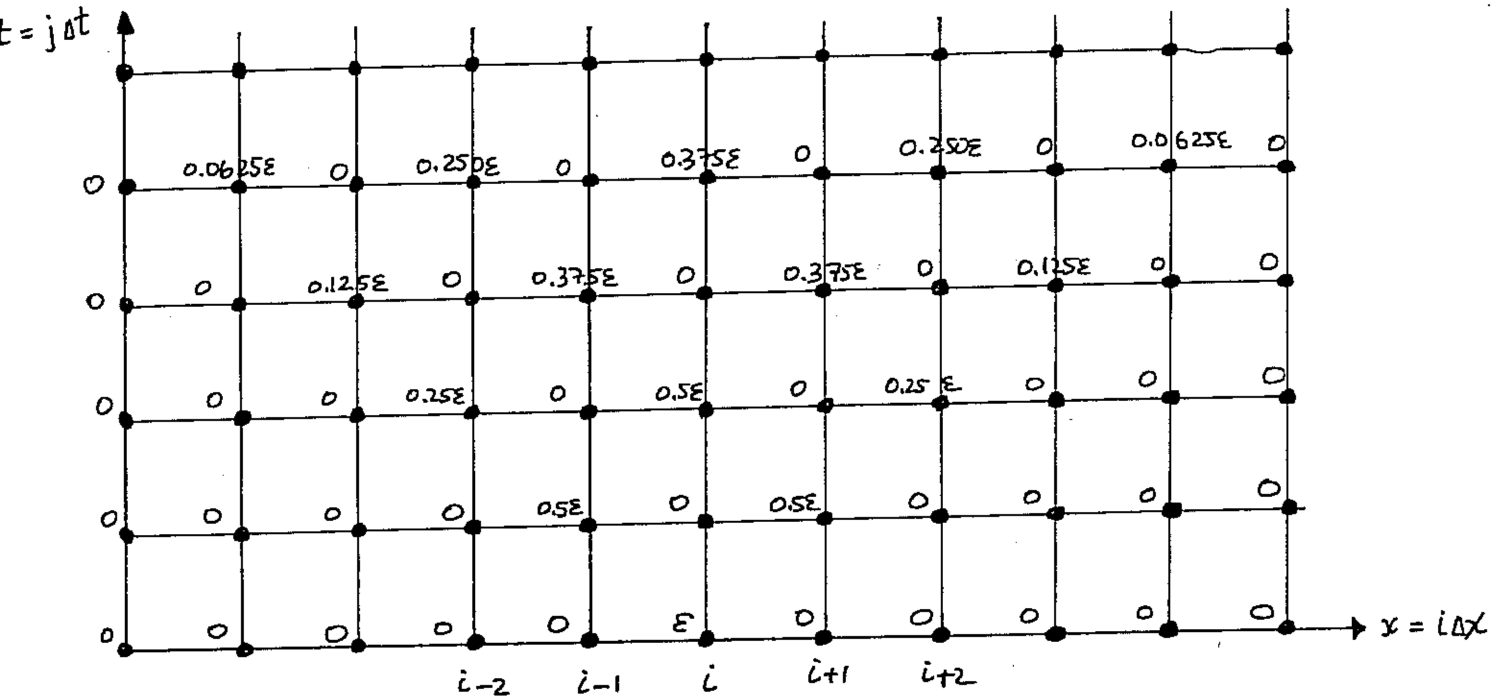
results in the difference equation for this specific ratio of $\rho = \dfrac{\Delta t D}{(\Delta x)^2}$:

$$u_{i,j+1} = \frac{1}{2} u_{i+1,j} + \frac{1}{2} u_{i-1,j}$$

This corresponds to the following molecule:



Let's look at the development of the error as the solution marches through time (using the molecule):



- Thus the high wave number solution is damped out by the difference equations!

Case 2

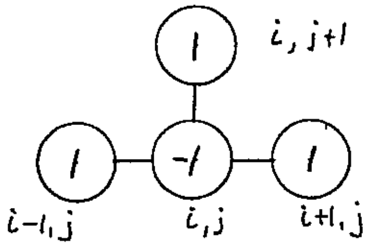
Letting,

$$\rho = 1$$

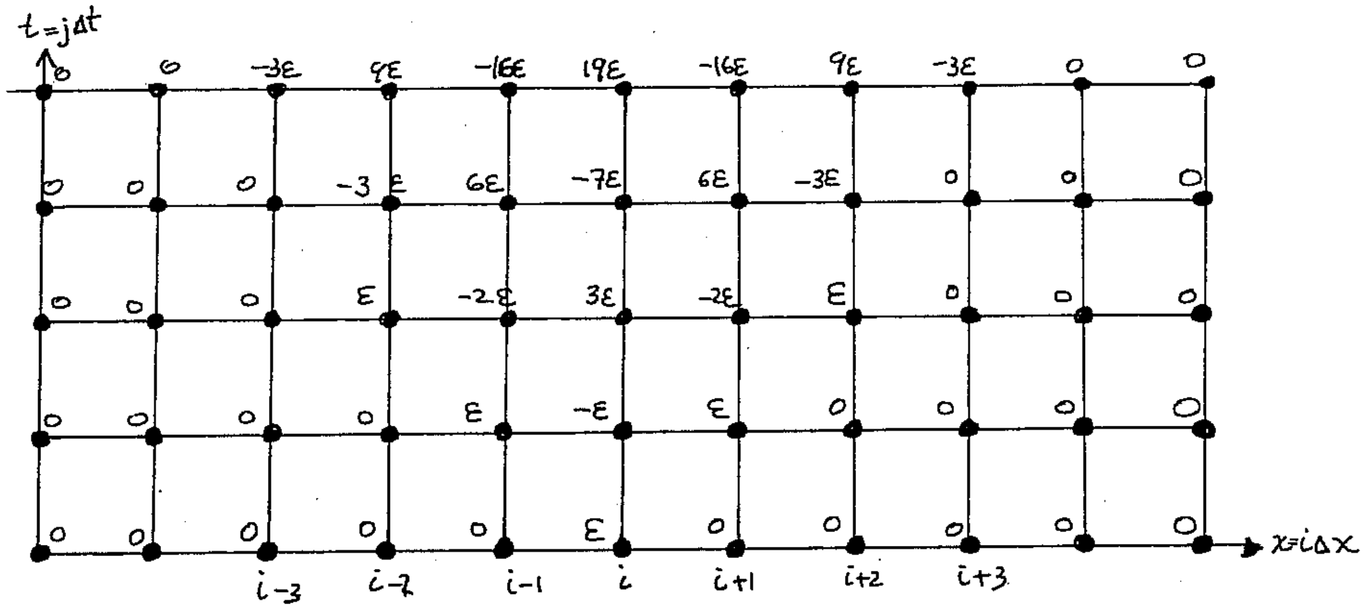
results in the difference equation for this specific ratio of $\rho = \dfrac{\Delta t D}{(\Delta x)^2}$:

$$u_{i,j+1} = u_{i+1,j} - u_{i,j} + u_{i-1,j}$$

This corresponds to the following molecule:



Let's examine how the solution propagates for this discretization.



- This solution is unstable. ***The amplitude of the solution grows in time.***
- In addition, note the oscillations in sign in both space and time. The oscillatory $2 \cdot \Delta x$ and $2 \cdot \Delta t$ behavior is often associated with instability.

Add a long wave example to show that solution will be stable

Notes on Instability

- Although we can have unstable growth for all wavelength components of a discrete solution (both short and long), typically the shorter wavelength components of a discrete solution experience the most rapid unstable growth.
- Eliminating the short wavelength energy from a discrete solution by increasing the level of resolution (i.e. the number of nodes in space and/or time) does not solve the problem of unstable growth of short wavelengths, if it exists, since roundoff always causes the existence of energy in the short wavelength range regardless of the level of resolution used!
- We desire bounds on instability and therefore we require a more formal analysis technique than the Heuristic approach.

Stability Analysis by Fourier Series Method (Von Neumann's Method)

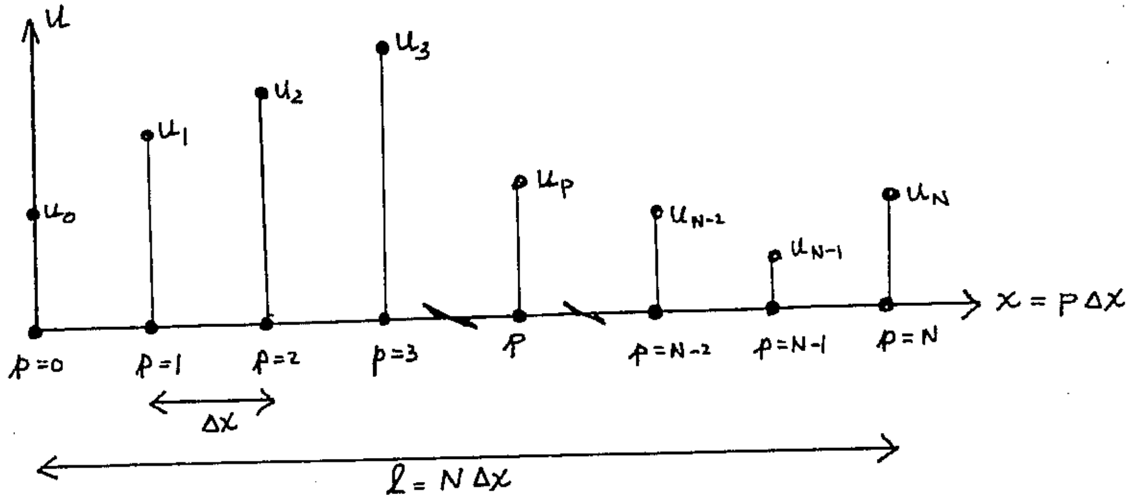
- Easy to use
- Not as rigorous as other methods since it neglects b.c.'s.
- ***Notational change:*** $i \rightarrow p$ ***(spatial index),*** $j \rightarrow q$ ***(time index)***

Step 1: Express the initial condition (at $t = 0$) in terms of a complex Fourier Series in space. Let:

$$u(x) = \sum_{n=0}^{N} A_n e^{in\pi x/l} \text{ where } i = \sqrt{-1}$$

The discrete solution at a given pivotal (or nodal) location is:

$$u(p\Delta x) = u_p \text{ for spatial points } p = 0, 1, 2, \ldots, N$$



This results in $N+1$ equations (1 at each pivotal point or node). This then allows us to determine the amplitude A_n at each pivotal point.

Letting $x = p\Delta x$ and $l = N\Delta x$

$$e^{in\pi x/l} = e^{in\pi p\Delta x/N\Delta x} = e^{i\beta_n p\Delta x}$$

where

$$\beta_n = \frac{n\pi}{N\Delta x}$$ = wave~~length~~number of wave component n

and the relationship between wavelength and wavenumber is defined as:

$$\beta_n = \frac{2\pi}{\lambda_n}$$

where

λ_n = wavelength of wave component n

With these considerations:

$$u_p = \sum_{n=0}^{N} A_n e^{i\beta_n p\Delta x}$$

- These $N + 1$ equations allow the unique determination of unknown coefficients $A_o, A_1, A_2, \ldots A_N$.
- The point is that we can exactly represent the i.c. and/or discrete solution at all the nodes by a Fourier series!
- ***Hence the i.c. (at time = 0) has been expressed as a Fourier Series!*** Therefore any i.c. or discrete solution can be expressed as a Fourier series (sines/cosines or in equivalent complex exponential form.)
- Since the FD equations are linear, separate solutions can be added. ***Thus we need only consider the propagation of the solution due to a single generic term in the Fourier series.***
- A wide range of wavelengths can exist in any i.c. and discrete solution.

  - The longest wavelength coincides with $n = 0$.

$$\beta_n = \frac{n\pi}{N\Delta x}$$

$$\Rightarrow$$

$$\beta_0 = 0 \quad \Rightarrow \quad \lambda_0 = \frac{2\pi}{0} \quad \Rightarrow \quad \lambda_0 = \infty$$

  - The shortest wavelength coincides with $n = N$.

$$\beta_N = \frac{N\pi}{N\Delta x}$$

$$\Rightarrow$$

$$\beta_N = \frac{\pi}{\Delta x} \quad \Rightarrow \quad \lambda_N = \frac{2\pi}{(\pi/\Delta x)} \quad \Rightarrow \quad \lambda_N = 2 \cdot \Delta x$$

- All other Fourier components of the i.c. and discrete solution range between these two extremes.

Due to the linearity of the difference equations, we consider only one component in the series. We note that the coefficient A_n is a constant which can now be neglected (since we're only considering one generic component).

Thus the a generic wavelength component representing a portion of the discrete i.c. can now be expressed as:

$$u_p = e^{i\beta p \Delta x}$$

<u>Step 2</u>: We can now investigate the development of the initial condition into the time varying discrete solution as t increases. We note that the solution to the FD equation must reduce to $u_p = e^{i\beta p \Delta x}$ when $t = q\Delta t = 0$ (i.e. the time varying solution must also satisfy the i.c.).

Let's assume that the solution of the FD equation may be written in separable form:

$$u_{p,q} = e^{i\beta x}e^{\alpha t}$$

$$\Rightarrow$$

$$u_{p,q} = e^{i\beta p \Delta x}e^{\alpha q \Delta t}$$

$$\Rightarrow$$

$$u_{p,q} = e^{i\beta p \Delta x}\xi^q$$

where

$$\xi = e^{\alpha \Delta t} = \textit{ the amplification factor}$$

$$\alpha = \text{in general a complex constant}$$

- We note that $u_{p,q}$ reduces to $u_p = e^{i\beta p \Delta x}$ when $q = 0$ (i.e. when $t = 0$).
- We note that the amplification factor equals the ratio of the solution at consecutive time steps. Thus:

$$\frac{u_{p,q+1}}{u_{p,q}} = \xi$$

- ***The generic discrete solution component with wavenumber β_n will not increase as t increases if the Von Neumann Condition for the amplification factor is satisfied:***

$$|\xi| \le 1 \text{ or } -1 \le \xi \le 1$$

- We note the importance of avoiding any amplification in any solution component as a solution time marches along! Even the slightest amplification (i.e. $|\xi|$ slightly greater than unity) will eventually lead to unbounded growth.
- $|\xi| \le 1$ is a necessary and sufficient condition for stability for 2 level time difference schemes (but it is not always sufficient for 3 or more level equations although it is always necessary!)
- The Von Neumann method or Fourier analysis method described strictly applies only to linear difference equations with constant coefficients and strictly speak-

ing only to i.v.p.'s with periodic initial data. In practice the method gives useful results even when application is not fully justified!

Lecture No. 7

Example Applications of Fourier Analysis

Example 1: Apply Fourier analysis to the explicit solution of the diffusion equation.

$$u_{p,q+1} = \rho u_{p+1,q} + (1-2\rho)\, u_{p,q} + \rho u_{p-1,q}$$

$$\text{where } \rho = \frac{\Delta t D}{\Delta x^2}$$

Since the difference equation is linear, we need only consider one generic Fourier component (which can represent any wavelength ranging from the shortest resolvable wavelength, $\lambda = 2 \cdot \Delta x$, to the longest wavelength, $\lambda = \infty$. Therefore we substitute the generic solution component of the form:

$$u_{p,q} = e^{i\beta p \Delta x}\xi^q$$

into the difference equation. This leads to:

$$e^{i\beta p\Delta x}\xi^{q+1} = \rho e^{i\beta(p+1)\Delta x}\xi^q + (1-2\rho)\, e^{i\beta p\Delta x}\xi^q + \rho e^{i\beta(p-1)\Delta x}\xi^q$$

$$\Rightarrow$$

$$\xi = \rho e^{i\beta\Delta x} + (1-2\rho) + \rho e^{-i\beta\Delta x}$$

$$\Rightarrow$$

$$\xi = (1-2\rho) + \rho\,(e^{i\beta\Delta x} + e^{-i\beta\Delta x})$$

However since:

$$e^{i\beta\Delta x} + e^{-i\beta\Delta x} = 2\cos\beta\Delta x$$

and

$$\cos\beta\Delta x = 1 - 2\sin^2\left(\frac{\beta\Delta x}{2}\right)$$

Substituting into the relationship for the amplification factor:

$$\xi = (1-2\rho) + 2\rho\left(1 - 2\sin^2\frac{\beta\Delta x}{2}\right)$$

$$\Rightarrow$$

$$\xi = 1 - 4\rho\sin^2\frac{\beta\Delta x}{2}$$

- ξ equals the amplification factor for the explicit solution to the diffusion equation.

Now examine what requirements must be satisfied such that $|\xi| \le 1$ for all wavenumbers β (which can range between $0 \to \frac{\pi}{\Delta x}$). Thus we require that:

$$-1 \le 1 - 4\rho\sin^2\frac{\beta\Delta x}{2} \le 1$$

- The upper bound is automatically satisfied since:

$$\rho = \frac{\Delta t D}{(\Delta x)^2} > 0 \text{ and } \sin^2 \ge 0$$

- For stability then

$$-1 \leq 1 - 4\rho \sin^2\left(\frac{\beta \Delta x}{2}\right)$$

$$\Rightarrow$$

$$\rho \leq \frac{1}{2\sin^2\left(\frac{\beta \Delta x}{2}\right)}$$

- The smallest wavenumber corresponds to $\beta = 0$ ($\lambda = \infty$ is the longest possible wavelength component of the solution) which leads to:

$$\rho \leq \infty$$

- Thus the long wavelength components do not appear to have any restrictions on $(\rho = (\Delta t D)/(\Delta x^2))$ ~~$\leq 1/2$~~ ***in order to avoid unstable growth (i.e. to keep*** $|\xi| \leq 1$***).***

- The highest wavenumber corresponds to $\beta = \pi/(\Delta x)$ ($\lambda = 2 \cdot \Delta x$ is the shortest possible wavelength component of the solution) which leads to:

$$\rho \leq \frac{1}{2\sin^2\left(\frac{\pi}{2}\right)} \quad \Rightarrow \quad \rho \leq \frac{1}{2}$$

- Thus the shortest wavelength components appear to restrict $(\rho = (\Delta t D)/(\Delta x^2)) \leq 1/2$ ***in order to avoid unstable growth of the short wavelength components (i.e. to keep*** $|\xi| \leq 1$***).***

- ***We must always select the most restrictive stability condition since WE ALWAYS HAVE high wavenumber components in our discrete solution due to either poorly resolved i.c.'s, roundoff error or even nonlinear transfer of energy to the high wavenumber range.***

- Thus in general for stability we <u>must</u> have

$$0 < \rho \leq \frac{1}{2}$$

- Note that for $\rho = 0$, we have $\Delta t = 0$. Thus the lower limit of the stability condition is automatically satisfied.

- An alternative analysis would simply examine the case which corresponds to the largest value of $\sin^2\left(\frac{\beta \Delta x}{2}\right)$ (which equals 1 and corresponds to $\beta \Delta x = \pi$). This leads to the most restrictive condition on ρ. Therefore:

$$\rho \leq \frac{1}{2} \text{ for stability}$$

- Note that we observed this stability condition to be true when applying the Heuristic approach!

Example 2: Examine the stability conditions for the weighted implicit/explicit approximation to the diffusion equation.

$$u_{p,q+1} - u_{p,q} = \frac{\Delta t D}{(\Delta x)^2} (\theta (u_{p+1,q+1} - 2u_{p,q+1} + u_{p-1,q+1})$$

$$+ (1-\theta) (u_{p+1,q} - 2u_{p,q} + u_{p-1,q}))$$

$$\text{where } \rho = \frac{\Delta t D}{(\Delta x)^2}$$

Substitute;

$$u_{p,q} = e^{i\beta p \Delta x} \xi^q$$

into the FD approximation. This leads to:

$$e^{i\beta p \Delta x} \xi^{q+1} (1 + 2\rho\theta) = e^{i\beta (p+1) \Delta x} \xi^{q+1} \rho\theta + e^{i\beta (p-1) \Delta x} \xi^{q+1} \rho\theta$$

$$+ e^{i\beta (p+1) \Delta x} \xi^q \rho (1-\theta) + e^{i\beta p \Delta x} \xi^q (1 - 2\rho + 2\theta\rho) + e^{i\beta (p-1) \Delta x} \xi^q \rho (1-\theta)$$

Now divide through by $e^{i\beta p \Delta x}$ and ξ^q and re-arrange:

$$\xi [1 + 2\rho\theta - \rho\theta (e^{i\beta \Delta x} + e^{-i\beta \Delta x})] =$$ ⊗

$$\rho (1-\theta) [e^{i\beta \Delta x} + e^{-i\beta \Delta x}] + (1 - 2\rho + 2\theta\rho)$$

However:

$$e^{i\beta\Delta x} + e^{-i\beta\Delta x} = 2 - 4\sin^2(\frac{\beta\Delta x}{2})$$

Substitute in and re-arrange:

$$\xi = \frac{1 - 4\rho(1-\theta)\sin^2(\frac{\beta\Delta x}{2})}{1 + 4\rho\theta\sin^2(\frac{\beta\Delta x}{2})}$$

For stable decay and thus stability, we must have:

$$|\xi| \le 1 \text{ or } -1 \le \xi \le 1$$

Again since $0 \le \theta \le 1$, $\rho > 0$ and $\sin^2(\frac{\beta\Delta x}{2}) > 0$, the upper bound is automatically satisfied. Thus we must satisfy:

$$-1 \le \frac{1 - 4\rho(1-\theta)\sin^2(\frac{\beta\Delta x}{2})}{1 + 4\rho\theta\sin^2(\frac{\beta\Delta x}{2})}$$

$$\Rightarrow$$

$$-1 - 4\rho\theta\sin^2(\frac{\beta\Delta x}{2}) \le 1 - 4\rho\sin^2(\frac{\beta\Delta x}{2}) + 4\rho\theta\sin^2(\frac{\beta\Delta x}{2})$$

Assume that $-4+8\theta < 0 \Rightarrow \theta < 0.5$ then must change direction of equality.

$$\Rightarrow$$

$-2 \le -4\rho\sin^2 + 8\rho\theta\sin^2$

$-2 \le (-4\rho + 8\rho\theta)\sin^2$

$-2 \le (-4+8\theta)\rho\sin^2$

$-4+8\theta < 0$

$\frac{-2}{(-4+8\theta)\sin^2} \ge \rho$

$$\rho \le \frac{-2}{(-4+8\theta)\sin^2(\frac{\beta\Delta x}{2})}$$

$$\Rightarrow$$

Alternatively if $-4+8\theta > 0 \Rightarrow \theta \ge 0.5$

$\rho \ge \frac{-2}{(-4+8\theta)\sin^2(\frac{\beta\Delta x}{2})} \Rightarrow$

$\rho \ge \frac{-2}{(\text{pos. no.})} \Rightarrow$ all that is necessary is that ρ is positive

$$\rho \leq \frac{1}{2\,(1-2\theta)\,\sin^2\left(\frac{\beta\Delta x}{2}\right)}$$

- Evaluating this expression over the range of $\beta\Delta x$ values:

  - The smallest wavenumber corresponds to $\beta = 0$ ($\lambda = \infty$ is the longest possible wavelength component of the solution) which leads to:

$$\rho \leq \infty$$

  - The highest wavenumber corresponds to $\beta = \pi/(\Delta x)$ ($\lambda = 2 \cdot \Delta x$ is the shortest possible wavelength component of the solution) which leads to:

$$\rho \leq \frac{1}{2\,(1-2\theta)}$$

  - Thus the long wavelength components do not have any restrictions on ρ ***in order to avoid unstable growth while the shortest wavelength components are restricted*** $\rho \doteq (\Delta t D)/(\Delta x^2) \leq 1/(2\,(1-2\theta))$ ***in order to avoid unstable growth.***

- Alternatively, we can simply evaluate the expression for ρ using the largest value of $\sin^2\left(\frac{\beta\Delta x}{2}\right)$ (which equals 1.0), leading to the most restrictive condition:

$$\rho \leq \frac{1}{2}\frac{1}{(1-2\theta)}$$

- Summary of stability conditions for the weighted implicit/explicit approximation to the diffusion equation:

  - classic explicit: $\theta = 0 \quad \rightarrow \quad$ Stability requires that $\rho \le \frac{1}{2}$.

  - $0 < \theta < \frac{1}{2} \quad \rightarrow \quad$ Stability requires that $\rho \le \frac{1}{2}\left(\frac{1}{1-2\theta}\right)$

  - Crank-Nicolson $\theta = \frac{1}{2} \quad \rightarrow \quad$ Stability requires that $\rho \le \infty$, i.e. the scheme is unconditionally stable.

  - $\frac{1}{2} < \theta \le 1 \quad \rightarrow \quad \rho \le \left[\frac{1}{2} \cdot \frac{1}{(1-2\theta)}\right]$ which is negative. However $\rho > 0$, and thus for this range of θ values the scheme is unconditionally stable. Alternatively we note that:

$$-1 \le \frac{1 - 4\rho\,(1-\theta)\sin^2\left(\frac{\beta\Delta x}{2}\right)}{1 + 4\rho\theta\sin^2\left(\frac{\beta\Delta x}{2}\right)}$$

  is always satisfied.

Example 3: Let's examine the wave equation:

$$\frac{\partial^2 U}{\partial t^2} = \frac{\partial^2 U}{\partial x^2}$$

Each term is discretized using a central approximation:

$$\frac{1}{\Delta t^2}(u_{p,q+1} - 2u_{p,q} + u_{p,q-1}) = \frac{1}{\Delta x^2}(u_{p+1,q} - 2u_{p,q} + u_{p-1,q})$$

Since the FD equation is linear, we need only consider one generic wavelength component. Thus we substitute in $u_{p,q} = e^{i\beta p \Delta x}\xi^q$ and rearrange. This leads to:

$$\xi^2 - 2A\xi + 1 = 0$$

$$\text{where } A = 1 - 2r^2 \sin^2\left(\frac{\beta \Delta x}{2}\right)$$

$$\text{and } r = \frac{\Delta t}{\Delta x}$$

We now have 2 roots to the equation. We must satisfy both restrictions:

$$\xi_1 = A + (A^2 - 1)^{1/2} \quad \xi_2 = A - (A^2 - 1)^{1/2}$$

Thus we must satisfy $|\xi_1| \le 1$ as well as $|\xi_2| \le 1$.

- Since r, Δx and β are all real $\Rightarrow A \le 1$
- When $A < -1 \Rightarrow |\xi_2| > 1$ and the solution is unstable.
- When $-1 \le A \le 1 \Rightarrow A^2 \le 1$, in which case we can express:

$$\xi_1 = A + i(1 - A^2)^{1/2} \text{ and } \xi_2 = A - i(1 - A^2)^{1/2}$$

- Thus:

$$|\xi_1| = |\xi_2| = \{A^2 + (1-A^2)\}^{1/2} = 1$$

Thus it is necessary that $-1 \le A \le 1$ in order to have stability. This implies that:

$$-1 \le 1 - 2r^2 \sin^2\left(\frac{\beta \Delta x}{2}\right) \le 1$$

The upper limit will always be satisfied and thus we require only that:

$$-1 \le 1 - 2r^2 \sin^2\left(\frac{\beta \Delta x}{2}\right)$$

Evaluating this expression over the range of $\beta \Delta x$ values:

- The smallest wavenumber corresponds to $\beta = 0$ ($\lambda = \infty$ is the longest possible wavelength component of the solution) which leads to:

$$-1 \le 1$$

- The highest wavenumber corresponds to $\beta = \pi/(\Delta x)$ ($\lambda = 2 \cdot \Delta x$ is the shortest possible wavelength component of the solution) which leads to:

$$-1 \le 1 - 2r^2$$

$$\Rightarrow$$

$$r = \frac{\Delta t}{\Delta x} \le 1$$

- Thus the long wavelength components do not have any restrictions on r in order to avoid unstable growth while the shortest wavelength components are restricted to $r = \frac{\Delta t}{\Delta x} \le 1$ in order to avoid unstable growth.

Notes on Stability Analysis and Convergence

- Lax's equivalence theorem states that given a properly posed linear i.v.p. and a linear FD approximation that satisfies the consistency condition, then stability is a necessary and sufficient condition for convergence

- A numerical method is convergent if:

$$\|U_{p,q} - u_{p,q}\| \to 0 \text{ as } \Delta x \to 0 \text{ and } \Delta t \to 0$$

$$\| \ \| = \text{some norm}$$

$$U_{p,q} = \text{the exact value}$$

$$u_{p,q} = \text{the numerical value}$$

- Note that in practice we can never achieve $\|U_{p,q} - u_{p,q}\| \to 0$ on a computer due to roundoff problems.

- Fourier analysis does <u>not</u> consider the effect of b.c.'s but only of interior points. However b.c.'s and their treatment can be of critical importance to a numerical scheme. To consider the effect of b.c.'s, we must use matrix methods (the matrix includes the b.c.'s) and compute the eigenvalues of the discretization matrix.

- ***We can also "qualitatively" compare the accuracy and other properties of numerical schemes using Fourier methods.*** This will be explored in much more detail in a subsequent lecture.