(ch. 1)

Lecture 2: Consistency & Stability

Lax Equivalence Theorem states that consistency & stability = convergence; this has been the cornerstone of numerical methods for differential equations.

Consistency To understand consistency, let's look at the PDE

$$(2.1) \quad \frac{\partial \xi}{\partial t} + u \frac{\partial \xi}{\partial x} = 0 \qquad \text{(assume appropriate ICs \& BCs)}$$

If we approximate the derivatives using a general order approximation, we get:

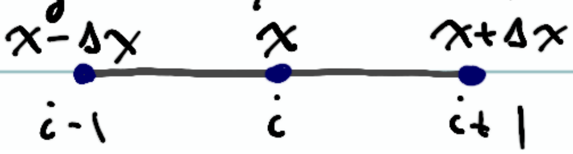
$$(2.2) \quad \sum_{k=-K}^{+K} \frac{\alpha_k \xi_i^{n+k}}{\Delta t} + \sum_{k=-N}^{+N} B_k \xi_{i+k}^{n} = 0$$

which yields an order $O(\Delta t^{2K}, \Delta x^{2N})$ method.

But let's simplify the discussion and assume $K = N = 1$ to get:

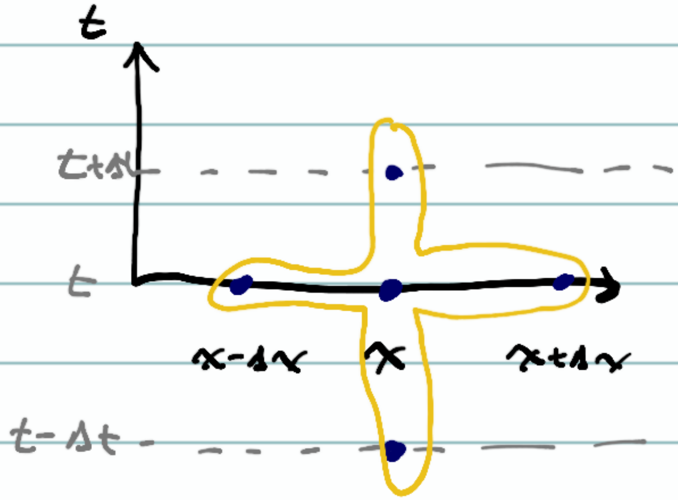
$$(2.3) \quad \frac{\alpha_{-1} \xi_i^{n-1} + \alpha_0 \xi_i^{n} + \alpha_1 \xi_i^{n+1}}{\Delta t} + \frac{u\, B_{i-1} \xi_{i-1}^{n} + B_0 \xi_i^{n} + B_1 \xi_{i+1}^{n}}{\Delta x} = 0$$

Taking a Taylor Series expansion about (x,t) s.t.



in the spatial stencil

which is similar for both time & space. The full stencil in the x-t diagram is:



Applying the Taylor Series yields:

$$
\begin{aligned}
g_i^n &= g(x,t) \equiv g \\
g_i^{n+1} &\equiv g(x,t+\Delta t) = g + \Delta t\, g_t + \frac{\Delta t^2}{2} g_{tt} + O(\Delta t^3) \\
g_i^{n-1} &\equiv g(x,t-\Delta t) = g - \Delta t\, g_t + \frac{\Delta t^2}{2} g_{tt} - O(\Delta t^3) \\
g_{i+1}^n &\equiv g(x+\Delta x,t) = g + \Delta x\, g_x + \frac{\Delta x^2}{2} g_{xx} + O(\Delta x^3) \\
g_{i-1}^n &= g(x-\Delta x,t) = g - \Delta x\, g_x + \frac{\Delta x^2}{2} g_{xx} - O(\Delta x^3)
\end{aligned}
\tag{2.4}
$$

For simplicity, pick $\alpha_{-1}=0$ (let's focus on the spatial discretization). Subbing (2.4) into (2.3) yields

$$
\begin{aligned}
&\frac{\alpha_1}{\Delta t}\left(g + \Delta t\, g_t + \frac{\Delta t^2}{2} g_{tt} + O(\Delta t^3)\right) + \frac{\alpha_0}{\Delta t} g \\
&+ \frac{u B_1}{\Delta x}\left(g + \Delta x\, g_x + \frac{\Delta x^2}{2} g_{xx} + O(\Delta x^3)\right) + \frac{u B_0}{\Delta x} g \\
&+ \frac{u B_{-1}}{\Delta x}\left(g - \Delta x\, g_x + \frac{\Delta x^2}{2} g_{xx} - O(\Delta x^3)\right) = 0
\end{aligned}
\tag{2.5}
$$

From the time-derivative terms we see that:

(TC1) q: $\frac{\alpha_1}{\Delta t} + \frac{\alpha_0}{\Delta t} = 0 \longrightarrow$ to eliminate q term

(TC2) $\frac{\partial q}{\partial t}$: $\frac{\alpha_1}{\Delta t} \Delta t = 1 \longrightarrow$ to keep q_t term

$\therefore \quad \alpha_1 = 1 \quad \& \quad \alpha_0 = -1 \quad \&$ we saw $\alpha_{-1} = 0$

From the space-derivative terms we note:

(SC1) q: $\beta_1 + \beta_0 + \beta_{-1} = 0 \longrightarrow$ to eliminate q term

(SC2) $\frac{\partial q}{\partial x}$: $\beta_1 - \beta_{-1} = 1 \longrightarrow$ to keep $u\frac{\partial q}{\partial x}$ term

2 eqs. w/ 3 unkn, ∴ is over-determined so we do the following: then β_0 is a free parameter. Pick $\beta_0 = 0$

$$\left(\begin{array}{l} \beta_1 + \beta_{-1} = -\beta_0 \\ \beta_1 - \beta_{-1} = 1 \end{array}\right)$$

adding rows 1 & 2 gives:
$2\beta_1 = 1 - \beta_0$ (SC3)

Subbing in the coefficients into Eq. (2.3) yields:

$$q_t + u q_x + \underbrace{u(\beta_1 + \beta_{-1})\frac{\Delta x}{2} q_{xx} + O(\Delta t, \Delta x^2)}_{\text{truncation error term}} = 0$$

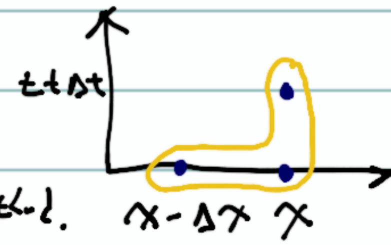
Let's 3 possible choices for our free parameter β_0.

Case 1 Let $\beta_0 = 1$ which, from SC3, $\beta_1 = 0$ & from SC2 $\beta_{-1} = -1$ that yields (from Eq. (2.3))

$$\frac{q_i^{n+1} - q_i^n}{\Delta t} + u \frac{q_i^n - q_{i-1}^n}{\Delta x} + u(0-1)\frac{\Delta x}{2} q_{xx} + O(\Delta t, \Delta x^2) = 0$$

or

$$\frac{q_i^{n+1} - q_i^n}{\Delta t} + u \frac{q_i^n - q_{i-1}^n}{\Delta x} + O(\Delta t, \Delta x) = 0$$

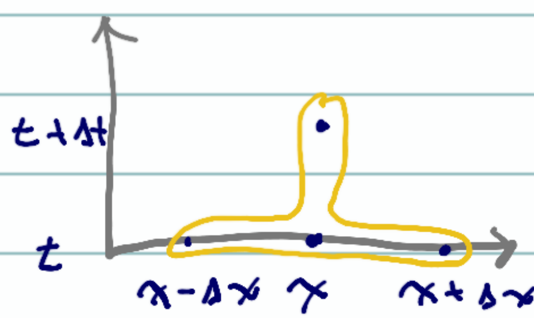


with the corresponding stencil which for $u>0$ is an upwinding method.

Case 2 Let $\beta_0 = 0$ which yields $\beta_1 = \frac{1}{2}$ (SC3) & $\beta_{-1} = -\frac{1}{2}$ (SC2) to get:

$$\frac{q_i^{n+1} - q_i^n}{\Delta t} + u \frac{q_{i+1}^n - q_{i-1}^n}{2\Delta x} + O(\Delta t, \Delta x^2) = 0$$

w/ the stencil

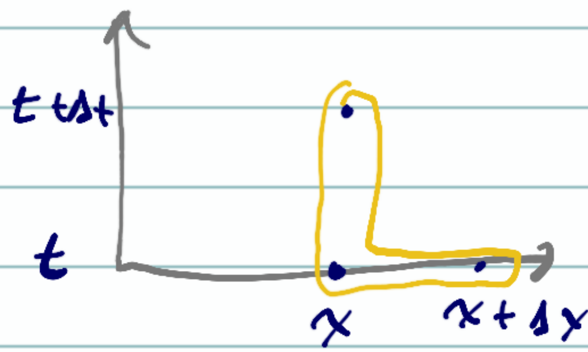


which is a centered method in space.

$\boxed{\text{Case 3}}$ Let $B_0 = -1$ which gives $B_1 = 1$ (SC1) & $B_{-1} = 0$ (SC2) to get:

$$\frac{f_i^{n+1} - f_i^{n}}{\Delta t} + u\,\frac{f_{i+1}^{n} - f_i^{n}}{\Delta x} + O(\Delta t, \Delta x) = 0$$

w/ Stencil



which for $u > 0$ is not an upwind method but rather a downwind method

Recap of Consistency we saw that for the 1D wave equation:

$$(2.1) \qquad \frac{\partial f}{\partial t} + u\,\frac{\partial f}{\partial x} = 0$$

we can discretize as

$$(2.2) \qquad \sum_{k=-K}^{+K} \frac{\alpha_k f_i^{n+k}}{\Delta t} + u \sum_{k=-N}^{+N} \frac{B_k f_{i+k}^{n}}{\Delta x} = 0$$

where "i" denotes a position in space & "n" in time.
Eq. (2.2) defines an $O(\leq 2K, \leq 2N)$ method.
For, e.g., $N = 1$ we have $k = -1, 0, +1$ or $2N+1$ points which gives an order $2N$ approximation

Applying a Taylor series, we obtained:

$$\xi_t + u\,\xi_x + u\,(B_1 + B_{-1})\,\frac{\Delta x}{2}\,\xi_{xx} + O(\Delta t, \Delta x^2) = 0$$

$\boxed{\text{Key Points}}$

1. as $(\Delta x, \Delta t) \rightarrow 0$ we recover the original continuous PDE $\rightarrow$ consistent
2. For $B_1 + B_{-1} \neq 0$ the method is $O(\Delta t, \Delta x)$
3. For $B_1 + B_{-1} = 0$ the method is $O(\Delta t, \Delta x^2)$

order $\leq 2K$, $\leq 2N$

Stability We will use Von Neumann stability analysis & matrix stability to check for the stability of methods. In this section we shall only discuss Von Neumann analysis.

In Fourier analysis we write the variable ξ_j as $\xi_j^n \equiv \xi(x_j, t^n) = \sum_{\ell=-N}^{+N} \tilde{\xi}_\ell^{(n)} e^{i k_\ell j \Delta x}$

with the following defs:

$\tilde{\xi}^{(n)} \rightarrow$ amplitude w/ exponent n

$e^{i k_\ell j \Delta x} \rightarrow$ frequencies with $i = \sqrt{-1}$

$k_\ell \rightarrow$ wave number

$\varphi_\ell = k_\ell \Delta x \rightarrow$ phase angle

For $\varphi \rightarrow 0$ we get low frequency (long waves) & for

$\varphi \to \pi$ we get high frequency (short waves)
N denotes the number of points in the domain.

To simplify our discussion, let us only consider a single Fourier mode & write it as follows:

$$\xi_j^n = \tilde{\xi}^{(n)} e^{ij\varphi}$$

Since $\tilde{\xi}$ is an amplitude then we must insist that $|\tilde{\xi}^{(n)}| < \infty$. For this to be satisfied require that $|\tilde{\xi}| \le 1$.

Let's now apply this to Cases 1, 2, & 3.

$\boxed{\text{Case 1}}$ For the upwind method $\dfrac{\xi_j^{n+1} - \xi_j^n}{\Delta t} + u\,\dfrac{\xi_j^n - \xi_{j-1}^n}{\Delta x} = 0$

we get:

(2.6) $$\xi_j^{n+1} = \xi_j^n - \frac{u\Delta t}{\Delta x}\left(\xi_j^n - \xi_{j-1}^n\right) \quad \& \quad \text{let } \sigma = \underbrace{\frac{u\Delta t}{\Delta x}}_{\text{Courant Number}}$$

Writing in terms of Fourier modes yields:

(2.7)
$$\xi_j^n = \tilde{\xi}^{(n)} e^{ij\varphi}$$
$$\xi_j^{n+1} = \tilde{\xi}^{(n+1)} e^{ij\varphi}$$
$$\xi_{j-1}^n = \tilde{\xi}^{(n)} e^{i(j-1)\varphi}$$

Subbing (2.7) into (2.6) yields:

(2.8) $$\tilde{\xi}^{(n+1)} e^{ij\varphi} = \tilde{\xi}^{(n)} e^{ij\varphi} - \sigma\left[\tilde{\xi}^{(n)} e^{ij\varphi} - \tilde{\xi}^{(n)} e^{i(j-1)\varphi}\right]$$

Dividing by $\tilde{\xi}^{(n)} e^{ij\varphi}$ gives:

$$\tilde{\xi} = 1 - \sigma\left(1 - e^{-i\varphi}\right)$$ → Euler's formula: $e^{\pm i\varphi} = \cos\varphi \pm i\sin\varphi$ gives:

$$\tilde{\xi} = 1 - \sigma\left[1 - (\cos\varphi - i\sin\varphi)\right] = 1 - \sigma + \sigma\cos\varphi - i\sigma\sin\varphi$$

when $Re(\tilde{\xi}) = (1-\sigma) + \sigma\cos\varphi$ & $Im(\tilde{\xi}) = -\sigma\sin\varphi$ that defines a circle of radius σ w/ center $(1-\sigma, 0)$ in the complex plane. Recall that we need $|\tilde{\xi}| \leq 1$ which gives:

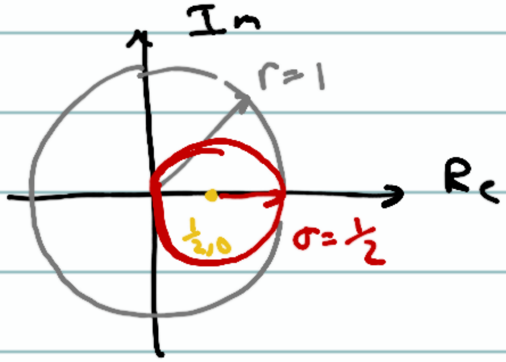
$$\tilde{\xi}\tilde{\xi}^* = \left[(1-\sigma) + \sigma\cos\varphi\right]^2 + (\sigma\sin)^2$$

Note that $|\tilde{\xi}| = (\tilde{\xi}\tilde{\xi}^*)^{\frac{1}{2}}$

defines a circle of radius σ w/ center $(1-\sigma, 0)$ ≤ 1

which will only occur for $\sigma \leq 1$ (Conditionally Stable)

Here is a schematic:



Note that for $\sigma = 1$, the circle starts at $(0,0)$

$\boxed{\text{Case 2}}$ For the centered method: $$\frac{\xi_j^{n+1} - \xi_j^n}{\Delta t} + u\frac{\xi_{j+1}^n - \xi_{j-1}^n}{2\Delta x} = 0$$

we require the following Fourier representation

$$\xi_j^n = \tilde{\xi}^{(n)} e^{ij\varphi}$$
$$\xi_j^{n+1} = \tilde{\xi}^{(n+1)} e^{ij\varphi}$$
$$\xi_{j+1}^n = \tilde{\xi}^{(n)} e^{i(j+1)\varphi}$$
$$\xi_{j-1}^n = \tilde{\xi}^{(n)} e^{i(j-1)\varphi}$$

which we sub into the differencing stencil to get:

$$\tilde{\xi}^{(n+1)} e^{ij\varphi} = \tilde{\xi}^{(n)} e^{ij\varphi} - \frac{u\Delta t}{2\Delta x}\left[\tilde{\xi}^{(n)} e^{i(j+1)\varphi} - \tilde{\xi}^{(n)} e^{i(j-1)\varphi}\right]$$

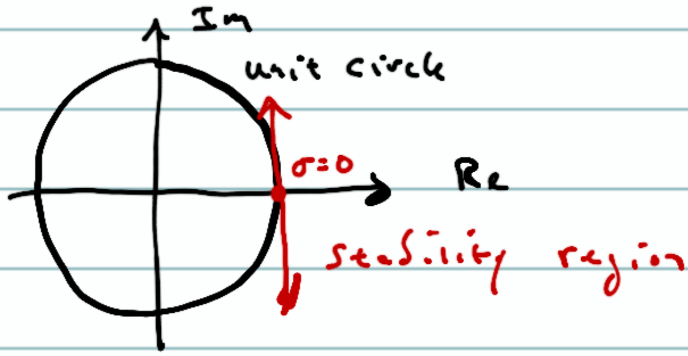
Dividing by $\xi_j^n = \tilde{\xi}^{(n)} e^{ij\varphi}$ gives:

$$\tilde{\xi} = 1 - \frac{\sigma}{2}\left[e^{i\varphi} - e^{-i\varphi}\right]$$ → using Euler's formula

$$\tilde{\xi} = 1 - \frac{\sigma}{2}\left[(\cancel{\cos\varphi} + i\sin\varphi) - (\cancel{\cos\varphi} - i\sin\varphi)\right]$$

$$= 1 - i\sigma\sin\varphi \qquad Re(\tilde{\xi}) = 1 \qquad Im(\tilde{\xi}) = -\sigma\sin\varphi$$

& $\tilde{\xi}\tilde{\xi}^* = 1 + \sigma^2\sin^2\varphi \leq 1$ only for $\sigma = 0$



$\boxed{\text{Case 3}}$ For the downwind method $\frac{\phi_j^{n+1}-\phi_j^{n}}{\Delta t} + u\frac{\phi_{j+1}^{n}-\phi_j^{n}}{\Delta x} = 0$

we get in Fourier space:

$$\tilde{\phi}^{(n+1)} e^{ij\varphi} = \tilde{\phi}^{(n)} e^{ij\varphi} - \frac{u\Delta t}{\Delta x}\left[\tilde{\phi}^{(n)} e^{i(j+1)\varphi} - \tilde{\phi}^{(n)} e^{ij\varphi}\right]$$

Dividing by $\tilde{\phi}^{(n)} e^{ij\varphi}$ gives:

$$\tilde{\phi} = 1 - \sigma\left[e^{i\varphi} - 1\right] \equiv 1 - \sigma\left[\cos\varphi + i\sin\varphi - 1\right]$$

$$\tilde{\phi} = 1 + \sigma - \sigma\cos\varphi - i\sigma\sin\varphi$$

$\mathrm{Re}(\tilde{\phi}) = 1 + \sigma - \sigma\cos\varphi \qquad \mathrm{Im}(\tilde{\phi}) = -\sigma\sin\varphi$

$$\tilde{\phi}\tilde{\phi}^{*} = (1 + \sigma - \sigma\cos\varphi)^2 + (\sigma\sin\varphi)^2 \le 1$$

which forms a circle of radius σ centered at $(1+\sigma, 0)$.

Here's a schematic: