Classification of PDEs

consider a general 2^{nd} Order PDE:

$$A u_{xx} + 2B u_{xy} + C u_{yy} = D$$

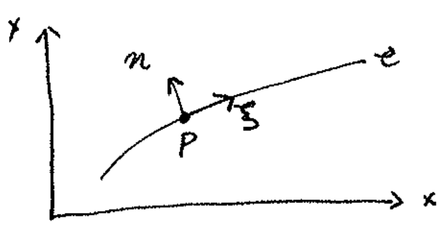
where $u(x,y)$ is the solution function, and A, B, C, D are smooth functions that depend on the independent variables x, y and on u, u_x, u_y (but not on the highest derivatives).

The equation is <u>nonlinear</u>, but it is <u>linear in its highest derivatives</u> → <u>quasi-linear</u>.

Laplace's eqn,	$u_{xx} + u_{yy} = 0$,	elliptic
"wave" eqn,	$u_{xx} - u_{yy} = 0$,	hyperbolic
heat eqn,	$u_{yy} = u_x$,	parabolic
burgers eqn,	$u_{yy} = u_x + u u_y$	
Tricomi eqn,	$u_{xx} - x u_{yy} = 0$	changing type

The solution behavior depends on A, B, C primarily but also on D and on any boundary or initial data.

<u>Classification</u>: examine the local behavior of solutions about a point P given some "Cauchy" data. Ask the question, can the solution be extended beyond P.

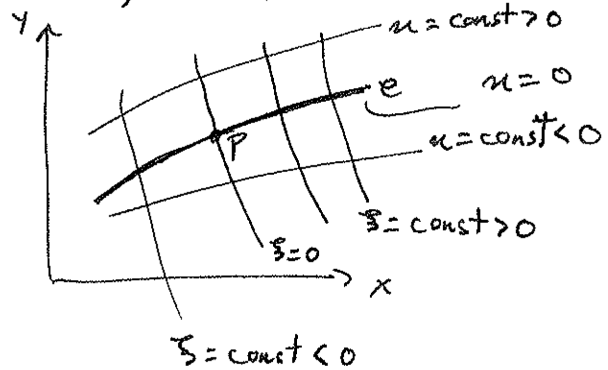


C is a smooth curve

let $\xi(x,y)$ measure position along C with $\xi = 0$ at the point P.

let $n(x,y)$ measure position away from C with $n = 0$ defining C

we can draw
contours, $\eta(x,y) = \text{const}$, $\xi(x,y) = \text{const}$.



we can define the curve $\mathcal{C}$ as

$\mathcal{C}: \eta(x,y) = 0$

assume that $\xi(x,y)$ and $\eta(x,y)$ are known about P.

Cauchy data: assume that u, u_y, u_x are given on $\mathcal{C}$ (about P).

the task is to use the data given on $\mathcal{C}$ and the PDE
to construct the solution in the neighborhood of P. If this
can be done for any $\mathcal{C}$ about P $\Rightarrow$ solution is analytic,
(ie has a Taylor series) and the equation is elliptic. If
on the other hand there are curves $\mathcal{C}$ for which the
solution can not be constructed off $\mathcal{C}$ then the
solution need not be analytic and the equation is
hyperbolic or parabolic

use a Taylor series

$$u(\xi,\eta) = \underbrace{u(\xi,0)}_{\text{given}} + \eta \underbrace{u_\eta(\xi,0)}_{\text{given}} + \underbrace{\frac{\eta^2}{2} u_{\eta\eta}(\xi,0)}_{\substack{\text{determine using data}\\ \text{on } \mathcal{C} \text{ and PDE}}} + \cdots$$

to compute $u_{\eta\eta}$, need $u_x, u_y, u_{xx}, u_{xy}, u_{yy}$ from chain rule

$\rightarrow u_\eta = u_x x_\eta + u_y y_\eta$

$\rightarrow u_{\eta\eta} = (u_{xx} x_\eta + u_{xy} y_\eta) x_\eta + u_x x_{\eta\eta} + (u_{yx} x_\eta + u_{yy} y_\eta) y_\eta + u_y y_{\eta\eta}$

Need to determine $u_x, u_y, u_{xx}, u_{xy}, u_{yy}$ along $\mathcal{C}$.

$u_x = u_\xi \xi_x + u_\eta \eta_x$
$u_y = u_\xi \xi_y + u_\eta \eta_y$

on $\mathcal{C}$, we know u_ξ and u_η therefore u_x and u_y are known on $\mathcal{C}$ from the given Cauchy problem

Now differentiate u_x, u_y along $\mathcal{C}$

$$\left.\begin{aligned}(u_x)_\xi &= u_{xx} x_\xi + u_{xy} y_\xi \\ (u_y)_\xi &= u_{yx} x_\xi + u_{yy} y_\xi\end{aligned}\right\}$$ the LHS is known from Cauchy data

use these two eqns with the PDE to determine u_{xx}, u_{xy}, and u_{yy} $\Rightarrow$ linear system

$$\begin{pmatrix} x_\xi & y_\xi & 0 \\ 0 & x_\xi & y_\xi \\ A & 2B & C \end{pmatrix}\begin{pmatrix} u_{xx} \\ u_{xy} \\ u_{yy} \end{pmatrix} = \begin{pmatrix} (u_x)_\xi \\ (u_y)_\xi \\ D \end{pmatrix}$$

The equation is solvable uniquely iff $\det(\text{matrix}) \neq 0$

$\rightarrow \quad A(y_\xi)^2 - 2B x_\xi y_\xi + C x_\xi^2 \neq 0 \quad (1)$

IF (1) holds for any smooth curve $\mathcal{C}$ through P, then we can construct Taylor series about P and therefore the solution is analytic $\Rightarrow$ ELLIPTIC CASE

Suppose $\mathcal{C}$ is such that $A y_\xi^2 - 2B x_\xi y_\xi + C x_\xi^2 = 0$ then $\mathcal{C}$ is called a characteristics.

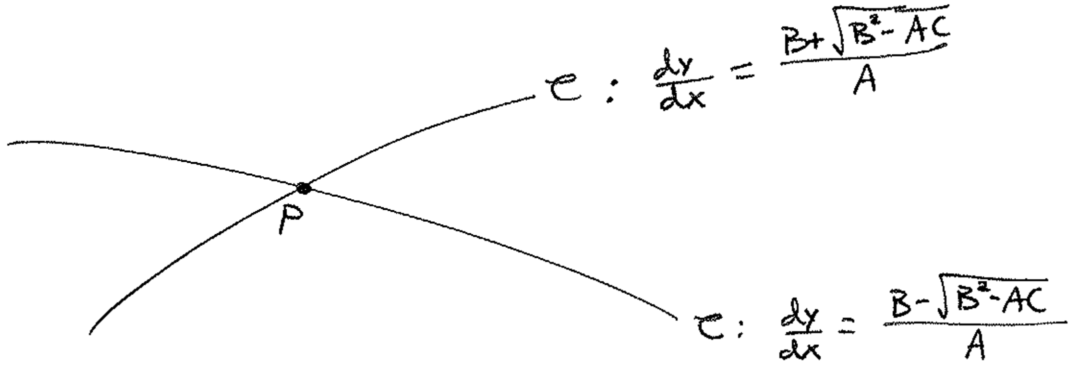
Suppose $A \neq 0$ and note that $\frac{y_\xi}{x_\xi} = \frac{dy}{dx}$ of $\mathcal{C}$ at P

$$A\left(\frac{dy}{dx}\right)^2 - 2B\frac{dy}{dx} + C = 0$$

$$\rightarrow \quad \frac{dy}{dx} = \frac{B \pm \sqrt{B^2 - AC}}{A}$$

→ there are 3 cases

1) $B^2 - AC \geqslant 0 \Rightarrow$ hyperbolic case



2) $B^2 - AC < 0 \Rightarrow$ elliptic case
no real characteristics, solutions are analytic

3) $B^2 - AC = 0 \Rightarrow$ parabolic case
coincident real characteristics

Characteristics

Consider the hyperbolic case and the linear system

$$\begin{pmatrix} x_\xi & y_\xi & 0 \\ 0 & x_\xi & y_\xi \\ A & B & C \end{pmatrix} \begin{pmatrix} u_{xx} \\ u_{xy} \\ u_{yy} \end{pmatrix} = \begin{pmatrix} u_{x\xi} \\ u_{y\xi} \\ D \end{pmatrix}$$

there are curves (ie characteristics) for which the determinant is zero. The system is singular so that solutions exist only if

$$\boxed{A y_\xi (u_x)_\xi + C x_\xi (u_y)_\xi = D x_\xi y_\xi}$$

we show this below ↓

$$(\alpha \;\; \beta \;\; \gamma) \begin{pmatrix} x_\xi & y_\xi & 0 \\ 0 & x_\xi & y_\xi \\ A & 2B & C \end{pmatrix} = 0$$

$$\Rightarrow \left(\alpha x_\xi + \gamma A, \; \alpha y_\xi + \beta x_\xi + 2\gamma B, \; \beta y_\xi + \gamma C \right) = 0$$

pick $\alpha = \frac{-\gamma A}{x_\xi}$, $\beta = \frac{-\gamma C}{y_\xi}$

Must have $(\alpha, \beta, \gamma)\begin{pmatrix} u_{x\xi} \\ u_{y\xi} \\ D \end{pmatrix} = 0$

$\rightarrow \quad \alpha u_{x\xi} + \beta u_{y\xi} + \gamma D = 0$

insert choices for α and β to obtain

$$-\frac{\gamma A}{x_\xi}(u_x)_\xi - \frac{\gamma C}{y_\xi}(u_y)_\xi + \gamma D = 0$$

$\rightarrow \quad A y_\xi (u_x)_\xi + C x_\xi (u_y)_\xi = D x_\xi y_\xi$ } holds along characteristics

Generally this equation is not solvable with simple analytic functions. There are some exceptions:

- <u>wave eqn</u>, $u_{xx} - u_{yy} = 0$, $A = 1$, $B = 0$, $C = -1$, $D = 0$

  $\rightarrow \quad \frac{dy}{dx} = \frac{B \pm \sqrt{B^2 - AC}}{A} = \pm 1 \quad \rightarrow \quad \boxed{\begin{array}{c} y = \pm x + \text{const} \\ \text{characteristics} \end{array}}$

  now examine eqn that holds along characteristics

  consider $y = x + \text{const}$

  $\rightarrow \quad y_\xi = x_\xi$

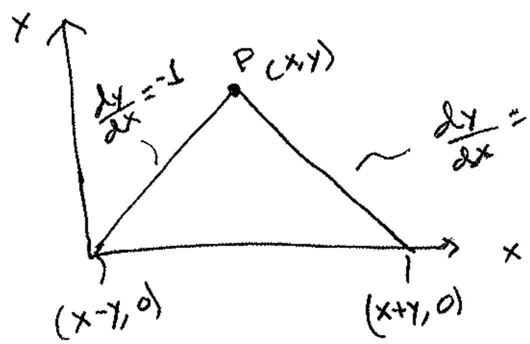
  $\rightarrow \quad y_\xi (u_x)_\xi - x_\xi (u_y)_\xi = 0 \quad \rightarrow \quad (u_x)_\xi - (u_y)_\xi = 0$

  $\rightarrow \quad u_x - u_y = f(\eta) = $ constant along each characteristic $y = x + \text{con}$

  Likewise, $y = -x + \text{const}$, $\Rightarrow$ $u_x + u_y = $ const.

  use characteristics and Riemann variables to construct so

  eg, $u_{xx} - u_{yy} = 0$, $|x| < \infty$, $y > 0$, $u(x,0) = f(x)$, $u_y(x,0) = g($



$$(u_x + u_y)\big|_p = f'(x+y) + g(x+y)$$

$$(u_x - u_y)\big|_p = f'(x-y) - g(x-y)$$

add and subtract

$$u_x(x,y) = \tfrac{1}{2} f'(x-y) + \tfrac{1}{2} f'(x+y) + \tfrac{1}{2} g(x+y) - \tfrac{1}{2} g(x-y)$$

$$u_y(x,y) = \tfrac{1}{2} f'(x+y) - \tfrac{1}{2} f'(x-y) + \tfrac{1}{2} g(x+y) + \tfrac{1}{2} g(x-y)$$

integrate with respect to x

$$u(x,y) = \tfrac{1}{2} f(x+y) + \tfrac{1}{2} f(x-y) + \tfrac{1}{2}\int_{x-y}^{x+y} g(s)\,ds + h(y)$$

differentiate with respect to y, compare to $u_y(x,y)$ from above, and conclude that $h'(y) = 0$

$\rightarrow h(y) = 0$ and $u(x,0)$ shows that $h(y) = 0$

to obtain d'Alembert's solution

Previously

considered the eqn $Au_{xx} + 2Bu_{xy} + Cu_{yy} = D$

where A, B, C, D depend on u, u_x, u_y.

considered local behavior near a given curve C

where u and its normal derivative is specified.

---

hyperbolic case, $B^2 - AC > 0$, 2 real characteristics

elliptic case, $B^2 - AC < 0$, no real characteristics

parabolic case, $B^2 - AC = 0$, 1 real characteristic

We would like to show that equation types are invariant under coordinate transformation.

Introduce new independent variables (s,t), ie $s = s(x,y)$, $t = t(x,y)$ with Jacobian non-zero (so the mapping isn't singular) and bounded.

$$J \equiv \frac{\partial(s,t)}{\partial(x,y)} = s_x t_y - s_y t_x \neq 0 \quad , \text{(and bounded)}$$

The chain rule gives

$$\boxed{(2) \quad \alpha u_{ss} + 2\beta u_{st} + \gamma u_{tt} = \delta}$$

where
$$\alpha = A s_x^2 + 2B s_x s_y + C s_y^2$$
$$\beta = A s_x t_x + B(s_x t_y + s_y t_x) + C s_y t_y$$
$$\gamma = A t_x^2 + 2B t_x t_y + C t_y^2$$

You can show that

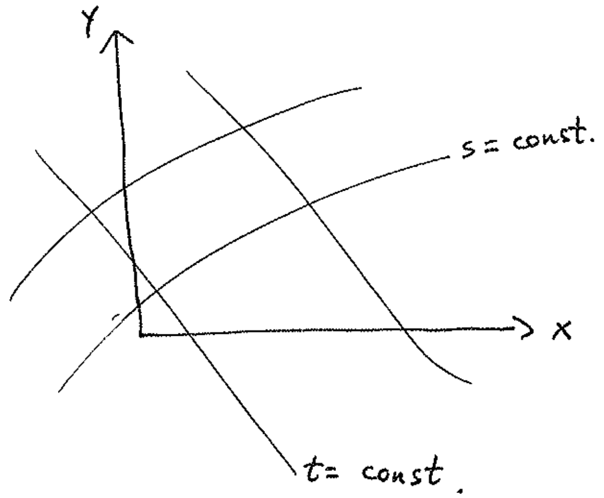
$$\beta^2 - \alpha\gamma = (B^2 - AC)\left(\frac{\partial(s,t)}{\partial(x,y)}\right)^2$$

$\Rightarrow$ the sign of the discrimanant is invariant $\Rightarrow$ the equation type is unchanged

Suggests that we could choose a coordinate system to bring ② to a certain canonical form.

Hyperbolic Case

Let $s(x,y) = $ const. be one family of characteristics and $t(x,y) = $ const. be the other family.



$$\Downarrow$$

eqn on s = const using t as parameter

$$A y_t^2 - 2B x_t y_t + C x_t^2 = 0$$

$$A y_s^2 - 2B x_s y_s + C x_s^2 = 0$$

eqn for characteristic t = const, using s as parameter

use $\quad S_x = \frac{1}{J} y_t \quad , \quad S_y = \frac{-1}{J} x_t \quad , \quad t_x = \frac{-1}{J} y_s \quad , \quad t_y = \frac{1}{J} x_s$

Recall,

$$\alpha = A s_x^2 + 2B s_x s_y + C s_y^2$$

$$\alpha = A\left(\frac{1}{J} y_t\right)^2 + 2B\left(\frac{1}{J} y_t\right)\left(\frac{-1}{J} x_t\right) + C\left(\frac{-1}{J} x_t\right)^2$$

$$\alpha = \frac{1}{J^2}\Big(\underbrace{A y_t^2 - 2B x_t y_t + C x_t^2}_{\text{vanishes}}\Big)$$

$$\alpha = 0$$

Likewise, can show that $\gamma = 0$

$\Rightarrow$ Equation ② reduces to the following hyperbolic canonical eqn

$$u_{st} = \delta$$

In the homogeneous problem where $\delta = 0$

$$u_{st} = 0 \quad \rightarrow \quad u = F(s) + G(t)$$

Another form: let $\hat{s} = s - t$
$\hat{t} = s + t$

then obtain $u_{\hat{s}\hat{s}} - u_{\hat{t}\hat{t}} = 0$, the other ~~characteristic~~ canonical form

Elliptic Case

no real characteristics, however use real and imaginary parts of the complex characteristics

suppose $(s + it) = \text{const.}$ is a complex characteristics

suppose that the characteristic is parametrized by a real variable ξ, ie

$$s(x(\xi), y(\xi)) + i\,t(x(\xi), y(\xi)) = \text{const.}$$

Differentiate:

$$(s_x + i t_x)\, x_\xi + (s_y + i t_y)\, y_\xi = 0$$

$$\rightarrow \quad y_\xi = -\frac{(s_x + i t_x)}{(s_y + i t_y)}\, x_\xi$$

this complex characteristic is defined by

$$A y_\xi^2 - 2 B x_\xi y_\xi + C x_\xi^2 = 0$$

Eliminate y_ξ:

$$A\left[-\frac{(s_x + i t_x)}{(s_y + i t_y)}\, x_\xi\right]^2 - 2 B x_\xi \left[\frac{-(s_x + i t_x)}{(s_y + i t_y)}\, x_\xi\right] + C x_\xi^2 = 0$$

After some algebra, the real part becomes:

$$\underbrace{A s_x^2 + 2 B s_x s_y + C s_y^2}_{\alpha} = \underbrace{A t_x^2 + 2 B t_x t_y + C t_y^2}_{\gamma}$$

$$\Rightarrow \quad \boxed{\alpha = \gamma \text{, in this coordinate system}}$$

the imaginary part tells us that $\beta = 0$ (in this coordinate system)

Therefore in this coordinate system, we have

$$\alpha u_{ss} + \gamma u_{tt} = \delta$$

with $\alpha = \gamma$, we have $u_{ss} + u_{tt} = \frac{\delta}{\alpha}$, Poisson's formula

the homogeneous eqn gives us $u_{ss} + u_{tt} = 0$, Laplace's eqn

Parabolic Case

Assume that $A \neq 0$ and take as our coordinates

$t =$ const is the characteristic

$s = x$

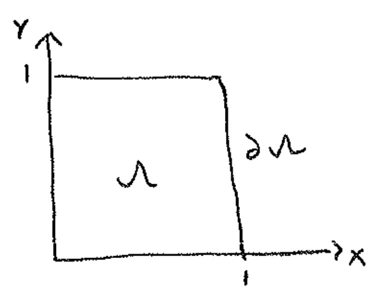
Can show that $\alpha = A \neq 0$, $\beta = \gamma = 0$ $\Rightarrow$ $\boxed{u_{ss} = \bar{\delta} = \frac{\delta}{A} \text{ canonical form}}$

Interesting case, $u_{ss} = u_t$, heat eqn

Boundary Conditions

Elliptic Case, the canonical form is Laplaces eqn.

Consider unit square as domain, and solve Laplace's eqn.



For a well-posed problem, need one condition on the whole perimeter. There are several forms:

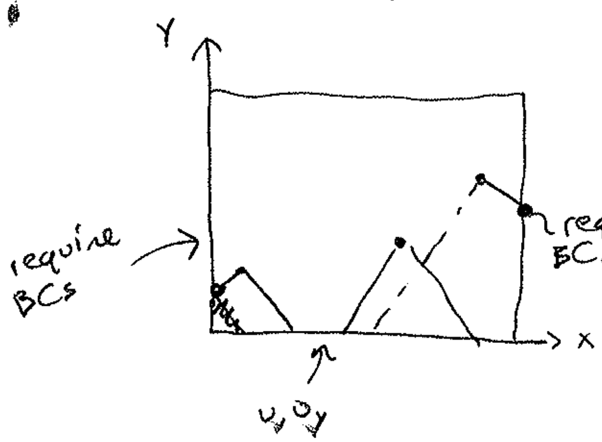
a) $u(x,y)$ given on $\partial\Omega$, (Dirichlet)

b) $\frac{\partial u}{\partial n} = \nabla u \cdot \hat{n}$ given on $\partial\Omega$, (Neumann)

c) $\alpha u + \beta \frac{\partial u}{\partial n}$ given on $\partial\Omega$, (Robin)

Hyperbolic Case (Boundary Conditions)

canonical form, $u_{xx} - u_{yy} = 0$



characteristics, $x+y = \text{const}$
$x-y = \text{const}$

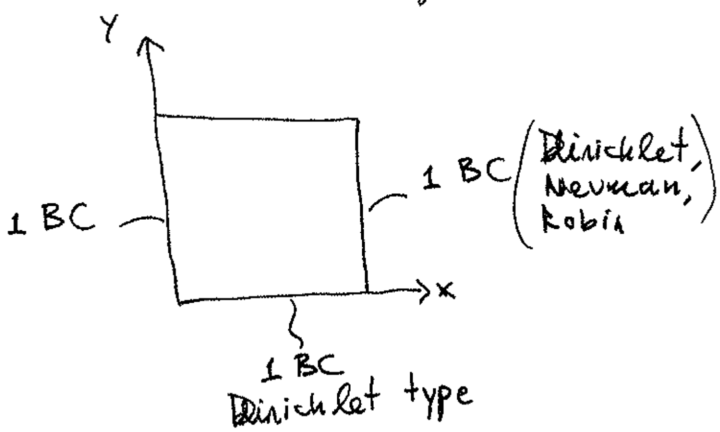
require two conditions on $y=0$

typically specify u, u_y given

domain of dependence / region of influence

Parabolic Case (BCs)

canonical form, $u_y = u_{xx}$, heat eqn



First Order Eqns

$$A u_x + B u_y = C$$

where A, B, C may depend on x, y, u (at most) $\Rightarrow$ quasi-linear

Always have one family of characteristics $\Rightarrow$ hyperbolic.

Suppose ξ is a parameter along the characteristic.

Set $\dfrac{dx}{d\xi} = A(x,y,u)$, $\dfrac{dy}{d\xi} = B(x,y,u)$

These two differential eqns define a path, where ξ measures position along the path. On the characteristic

$u(x,y) = u(x(\xi), y(\xi))$

$\rightarrow \dfrac{du}{d\xi} = u_x \dfrac{dx}{d\xi} + u_y \dfrac{dy}{d\xi} = A u_x + B u_y = C$

$\rightarrow$ Along the characteristic, $\dfrac{du}{d\xi} = C(x,y,u)$

Reduced the PDE to 3 ODEs.

First Order Systems

$$\underline{\underline{A}}\, \underline{u}_x + \underline{\underline{B}}\, \underline{u}_y = \underline{c}$$

The system may be elliptic, hyperbolic, parabolic, or a mix.

We'll talk about this later.

Finite Difference Methods for PDEs

Begin with a simple example involving <u>heat equation</u>.

$$u_t = \nu u_{xx} \quad , \quad 0 \le x \le 1 \, , \; t > 0 \quad , \nu = \text{constant}$$

$$u(x,0) = f(x) \quad , \quad u(0,t) = a(t) \, , \quad u(1,t) = b(t)$$

Assume $a(0) = f(0)$ and $b(0) = f(1)$ to enforce continuity, though not important due to diffusive nature of heat eqn.

Introduce a mesh

$$\left.\begin{array}{l} x_j = j\Delta x , \quad \Delta x = \frac{1}{N} \\ 0 = x_0 < x_1 < x_2 < \cdots < x_N = 1 \end{array}\right\} \text{spatial discretization}$$

define $v_j(t) \simeq u(x_j, t)$ ($v_j(t)$ approximates exact solution u at $x = x_j$ and time t)

Approximate derivatives

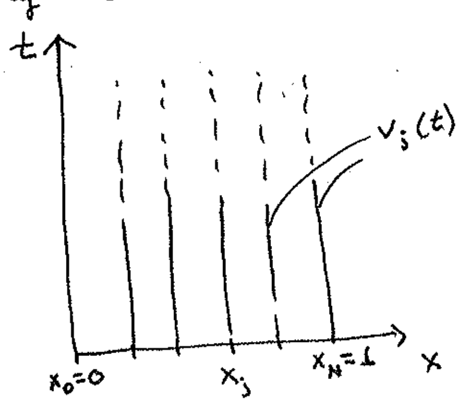
the usual thing for the heat eqn is to take a centered appx.

$$u_{xx}(x_j, t) \simeq \frac{1}{\Delta x^2}\left(v_{j-1}(t) - 2v_j(t) + v_{j+1}(t)\right) \quad - \; 2^{\text{nd}} \text{ order centered appx}$$

$$u_t(x_j, t) \simeq v_j'(t)$$

$$\Rightarrow \boxed{\begin{array}{l} v_j'(t) = \frac{\nu}{\Delta x^2}\left(v_{j-1}(t) - 2v_j(t) + v_{j+1}(t)\right) \quad , \quad j = 1, \ldots, N-1 \\ v_0(t) = a(t) \, , \quad v_N(t) = b(t) \quad , \; v_j(0) = f(x_j) \, , \; j = 1, .., N-1 \end{array}}$$

we have changed PDE into a bunch of ODEs
apply a method to solve ODEs numerically



"Method of lines" - discretize eqns in space first to obtain an ODE in time then use ODE solver to integrate ODEs

Solve the ODEs using Forward Euler;
ie replace $v'(t)$ with forward difference

$$v'(t) \simeq \frac{v_j(t+\Delta t) - v_j(t)}{\Delta t}, \quad \text{where } \Delta t \text{ is chosen}$$

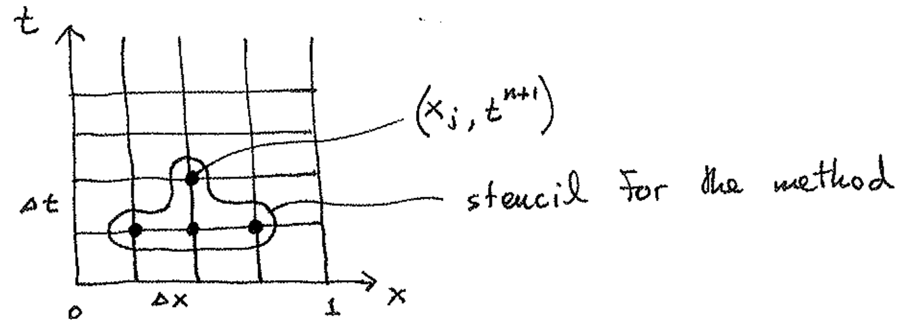
$$\rightarrow \quad \frac{v_j(t+\Delta t) - v_j(t)}{\Delta t} \simeq \nu \left[\frac{v_{j-1}(t) - 2v_j(t) + v_{j+1}(t)}{\Delta x^2}\right]$$

Suppose Δt is constant and $t_n = n\Delta t$ then
define $v_j^n \simeq v_j(t_n)$

$$\Rightarrow \boxed{\begin{array}{l} \dfrac{v_j^{n+1} - v_j^n}{\Delta t} = \nu \left[\dfrac{v_{j-1}(t) - 2v_j(t) + v_{j+1}(t)}{\Delta x^2}\right] \\ v_0^n = a(t_n), \quad v_N^n = b(t_n), \quad v_j^0 = f(x_j) \end{array}}$$

Finite difference method

we now have a full grid, (mesh)



you can solve the finite difference method explicity

here is the algorithm:

1) specify $N, \nu, \Delta t$
2) $v_j^0 = f(x_j)$ for $j = 0, \ldots, N$
3) for $n = 0, 1, \ldots, n_{Final}$
    4) $v_j^{n+1} = v_j^n + \frac{\Delta t \nu}{\Delta x^2}\left(v_{j-1}^n - 2v_j^n + v_{j+1}^n\right)$ for $j = 1, \ldots, N-1$
    5) $v_0^{n+1} = a(t_{n+1}), \quad v_N^{n+1} = b(t_{n+1})$

Example:

$$u_t = \nu u_{xx} \quad , \; 0 < x < 1, \; t > 0$$

$$u(x,0) = x + \sin \pi x + \sin 3\pi x$$

$$u(0,t) = 0 \quad , \quad u(1,t) = 1$$

Solve using $v_j^{n+1} = v_j^n + r\left(v_{j-1}^n - 2v_j + v_{j+1}^n\right)$, $r = \frac{\nu \Delta t}{\Delta x^2}$

Exact solution is

$$u(x,t) = x + e^{-\nu \pi^2 t} \sin \pi x + e^{-9\nu \pi^2 t} \sin 3\pi x$$

Solve numerically for $N = 20, 40, 80$ (spatial discretizations) and for various choices of Δt.

Issues

1) accuracy → error
2) instabilities? → choice for Δt
3) cost?

For $N = 20$, $r = 0.4$, $\max(\text{error}) = 9.84 \times 10^{-3}$
$N = 40$, $r = 0.4$, $\max(\text{error}) = 2.43 \times 10^{-3}$
$N = 80$, $r = 0.4$, $\max(\text{error}) = 6.05 \times 10^{-4}$

notice $\Delta x = \frac{1}{N}$ is halved and the error decreases by a factor of 4, which suggests that the the order decreases by Δx^2

$N = 40$, $r = 0.6$, ⇒ instability
→ limited by how big Δt can be

Consistency, Order of Accuracy

Previous example demonstrated that maximum error $\sim O(\Delta x^2)$ for $r = \frac{\nu \Delta t}{\Delta x^2}$ fixed less than some limiting value. Why?

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} = \nu \left(\frac{v_{j-1}^n - 2 v_j^n + v_{j+1}^n}{\Delta x^2} \right)$$

Introduce <u>difference operators</u>: (<u>linear operators</u>)

$\delta_{+x} v_j^n = v_{j+1}^n - v_j^n$, forward difference operator

$\delta_{-x} v_j^n = v_j^n - v_{j-1}^n$, backward difference operator

$\delta_{0x} v_j^n = v_{j+1}^n - \cancel{2} v_{j-1}^n$, centered difference operator

$$\begin{aligned} \delta_x^2 v_j^n &= \delta_{+x} \delta_{-x} v_j^n \\ &= v_{j-1}^n - 2 v_j^n + v_{j+1}^n , \end{aligned}$$ 2nd order centered difference operator

can also define in time

$$\delta_{+t} v_j^n = v_j^{n+1} - v_j^n$$

and so on

Using this notation, our finite difference method becomes

$$\boxed{\frac{1}{\Delta t} \delta_{+t} v_j^n = \frac{\nu}{\Delta x^2} \delta_x^2 v_j^n}$$

Introduce ~~an~~ the error grid function

$$e_j^n = v_j^n - u(x_j, t_n)$$

$\boxed{e_j^n = v_j^n - u_j^n}$, for notational convenience set $v_j^n = u_j^n + e_j^n$

substitute this into finite difference method

$$\rightarrow \frac{1}{\Delta t}\delta_{+t} u_j^n + \frac{1}{\Delta t}\delta_{+t} e_j^n = \frac{\nu}{\Delta x^2}\delta_x^2 u_j^n + \frac{\nu}{\Delta x^2}\delta_x^2 e_j^n$$

$$\rightarrow \boxed{\frac{1}{\Delta t}\delta_{+t} e_j^n - \frac{\nu}{\Delta x^2}\delta_x^2 e_j^n = -\left[\frac{1}{\Delta t}\delta_{+t} u_j^n - \frac{\nu}{\Delta x^2}\delta_x^2 u_j^n\right]} \Bigg\} \text{error equation}$$

$$e_j^0 = v_j^0 - u(x_j, 0) = 0$$

$$e_0^n = v_0^n - u(0, t_n) = 0$$

$$e_N^n = v_N^n - u(x_N, t_n) = 0$$

- Examine the RHS of the error equation

$$\frac{1}{\Delta t}\delta_{+t} u_j^n = \frac{u(x_j, t_n + \Delta t) - u(x_j, t_n)}{\Delta t}$$

$$= \frac{1}{\Delta t}\left[\Delta t\, u_t(x_j, t_n) + \frac{\Delta t^2}{2} u_{tt}(x_j, t_n) + \cdots\right]$$

$$= u_t(x_j, t_n) + \frac{\Delta t}{2} u_{tt}(x_j, t_n) + O(\Delta t^2)$$

$$\frac{1}{\Delta x^2}\delta_x^2 u_j^n = \frac{1}{\Delta x^2}\left[u(x_j - \Delta x, t_n) - 2u(x_j, t_n) + u(x_j + \Delta x, t_n)\right]$$

$$= \frac{1}{\Delta x^2}\left[\begin{aligned} &\cancel{u(x_j, t_n)} - \cancel{\Delta x\, u_x} + \frac{\Delta x^2}{2} u_{xx} - \cancel{\frac{\Delta x^3}{6} u_{xxx}} + \frac{\Delta x^4}{24} u_{xxxx} + \cdots \\ &\cancel{-2u} + \cancel{u} + \cancel{\Delta x\, u_x} + \frac{\Delta x^2}{2} u_{xx} + \cancel{\frac{\Delta x^3}{6} u_{xxx}} + \frac{\Delta x^4}{24} u_{xxxx} + \cdots \end{aligned}\right]$$

$$= u_{xx}(x_j, t_n) + \frac{\Delta x^2}{12} u_{xxxx}(x_j, t_n) + O(\Delta x^4)$$

The error equation becomes

$$\frac{1}{\Delta t}\delta_{+t} e_j^n - \frac{\nu}{\Delta x^2}\delta_x^2 e_j^n = -\left[\begin{array}{l} u_t + \frac{\Delta t}{2} u_{tt} + O(\Delta t^2) \\ -\nu\left(u_{xx} + \frac{\Delta x^2}{12} u_{xxxx} + O(\Delta x^4)\right) \end{array}\right]_{(x_j, t_n)}$$

since $u(x_j, t_n)$ solves the differential equation, what remains is the negative of the truncation error

$$\frac{1}{\Delta t}\delta_{+t} e_j^n - \frac{\nu}{\Delta x^2}\delta_x^2 e_j^n = -\underbrace{\left[\frac{\Delta t}{2} u_{tt} - \nu \frac{\Delta x^2}{12} u_{xxxx} + O(\Delta t^2, \Delta x^4)\right]}_{\text{local truncation error}}$$

<u>local truncation error</u>: amount by which the exact solution fails to satisfy the difference approximation

In our example, $r = \frac{\nu \Delta t}{\Delta x^2} = \text{const} \rightarrow \Delta t = \frac{r \Delta x^2}{\nu}$

Then the right hand side becomes

$$-\left[\frac{r\Delta x^2}{2\nu} u_{tt} - \frac{\nu \Delta x^2}{12} u_{xxxx} + \cdots\right]$$

$$= -\frac{\Delta x^2}{2\nu}\left[r u_{tt} - \frac{\nu^2}{6} u_{xxxx} + \cdots\right]$$

$$= O(\Delta x^2)$$

So if the solution is well-behaved then the size of e_j^n would be $O(\Delta x^2)$. This suggests that

$$\max_{j,n} |e_j^n| = O(\Delta x^2)$$

For our difference equation, the local truncation error is

$$\tau_j^n = \frac{1}{\Delta t}\,\delta_{+t}\,u_j^n - \frac{\nu}{\Delta x^2}\,\delta_x^2\,u_j^n$$

A difference method is said to be <u>consistent</u> if

$$\max_{j,n} |\tau_j^n| \to 0 \quad \text{as} \quad \Delta x, \Delta t \to 0$$

So for our example

$$\tau_j^n = \frac{\Delta t}{2} u_t - \frac{\nu \Delta x^2}{12} u_{xxxx} + \cdots$$

$$= O(\Delta t, \Delta x^2) \to 0 \quad \text{as} \quad \Delta t, \Delta x \to 0$$

The order of accuracy refers to the rate at which the truncation error τ_j^n goes to zero.

So for our example, the order of accuracy is first order in time and second order in space.

For our example, we had that

$$e_j^n \approx \Delta x^2\, E(x_j, t_n)$$

amplitude "shape of plot of error"

Go back to the error equation

$$\frac{1}{\Delta t}\,\delta_{+t}\,e_j^n - \frac{\nu}{\Delta x^2}\,\delta_x^2\,e_j^n = \Delta x^2 F_j^n + \cdots, \quad \text{with ICs \& BCs}$$

$$\to \frac{1}{\Delta t}\,\delta_{+t}\left(\frac{e_j^n}{\Delta x^2}\right) - \frac{\nu}{\Delta x^2}\,\delta_x^2\left(\frac{e_j^n}{\Delta x^2}\right) = F_j^n + \cdots$$

as $\Delta x \to 0$ for $r =$ fixed

$$E_t - \nu E_{xx} = F \quad + \text{ homogeneous ICs, BCs}$$

Computational Cost

number of Floating point operations required to compute v_j^n to a time $t_{Final} = O(1)$

number of flops needed per time step is $O(N)$

number of time steps is $O(1/\Delta t)$

if r is fixed then $O(1/\Delta t) = O(\Delta x^2) = O(1/N^2)$

$\rightarrow$ cost $= O(N^3)$

Best you could hope for is $O(N^2)$ ie one flop per grid point

A Neumann Problem for the Heat Equation

9/8/03

$u_t = \nu u_{xx}$, $0 < x < 1$, $t > 0$

IC: $u(x,0) = f(x)$

BC: $u_x(0,t) = 0$, $u_x(1,t) = 0$ - homogeneous Neumann BCs

Let us persue a Finite difference approximation of the problem.

Approximate the solution, $v_j(t) \simeq u(x_j, t)$

Require a finite difference eqn for $v_j(t)$:

$$\rightarrow v_j'(t) = \frac{\nu}{\Delta x^2} \delta_x^2 v_j(t)$$

, recall, $\delta_x^2 v_j(t) = v_{j-1} - 2v_j + v_{j+1}$

This equation is defined for some range of the subscript j.

In the Dirichlet case, did not have to solve at $j = 0, N$

The difficulty in the Neumann case is that $u(x,t)$ is not specified on the boundary.

For example at $j=1$,

$$v_1'(t) = \frac{\nu}{\Delta x^2}\left(v_2 - 2v_1 + v_0\right)$$

In the Dirichlet case, v_0 was given, but this is not the situation in the Neumann case.

We need to solve

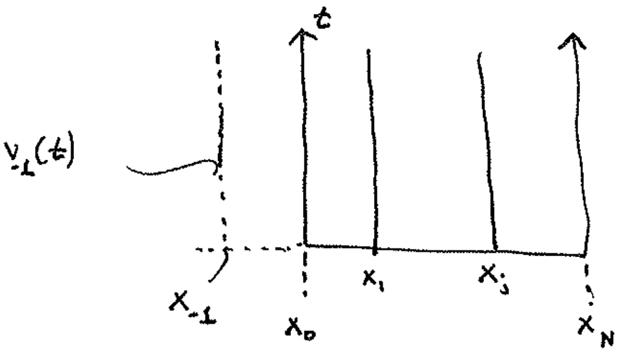
$$v_j'(t) = \frac{\nu}{\Delta x^2}\,\delta_x^2 v_j(t) \quad , \quad j=0,\dots,N$$

But now at $j=0$,

$$v_j'(0) = \frac{\nu}{\Delta x^2}\left(v_1 - 2v_0 + v_{-1}\right)$$

So we now have the negative subscript!

The Fix is to introduce a "ghost" line to handle the Neumann B.C.



consider the boundary condition

$$u_x(0,t) = 0$$

approximate the derivative using a finite difference equation

$$u_x\Big|_{x=0} \simeq \frac{1}{2\Delta x}\,\delta_{0x} v_0(t) \qquad \text{- centered difference, } \frac{v_1(t) - v_{-1}(t)}{2\Delta x}$$

the discrete BC is

$$\frac{1}{2\Delta x}\,\delta_{0x} v_0(t) = 0 \longrightarrow \delta_{0x} v_0(t) = 0 \longrightarrow \boxed{v_{-1}(t) = v_1(t)}$$

Therefore at $j=0$:

$$v_0'(t) = \frac{\nu}{\Delta x^2}\left(v_1 - 2v_0 + v_{-1}\right) = \frac{\nu}{\Delta x^2}\left(2v_1 - 2v_0\right)$$

Can choose to make this substitution into the FDE, in the program or just do the loop and use extra equation - just a matter of personal choice.

Similarly near $x=1$,

$$u_x(1,t) = 0 \rightarrow \frac{1}{2\Delta x}\delta_{0x} v_N = 0 \rightarrow \boxed{v_{N+1} = v_{N-1}}$$

The result of all this is a set of ODEs, solving for $v_j(t)$, $j=0,\dots,N$:

$$v_0' = \frac{\nu}{\Delta x^2}\left(2v_1 - 2v_0\right)$$

$$v_j' = \frac{\nu}{\Delta x^2}\left(v_{j+1} - 2v_j + v_{j-1}\right), \quad j = 1,\dots,N-1$$

$$v_N' = \frac{\nu}{\Delta x^2}\left(2v_{N-1} - 2v_N\right)$$

with ICs $v_j(0) = f(x_j)$, $j=0,\dots,N$

Now integrate the ODEs using method of your choice.

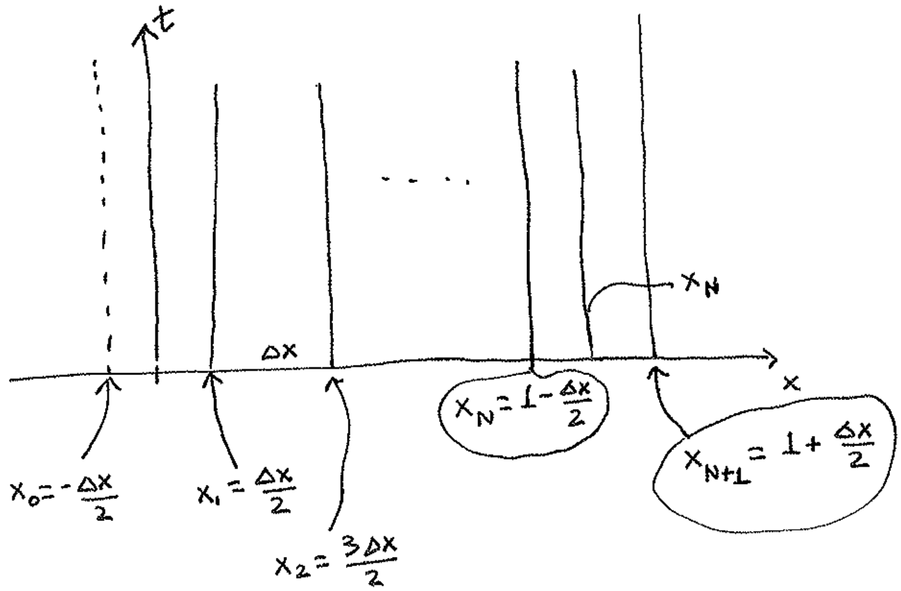
Alternate approach

$$u_t = \nu u_{xx}, \quad 0<x<1, \quad t>0$$

$$u(x,0) = f(x), \quad u_x(0,t) = u_x(1,t) = 0$$

Let us discuss the use of a staggered grid:

$$x_j = \left(j - \tfrac{1}{2}\right)\Delta x, \quad \Delta x = \frac{1}{N}$$



Use $v_j'(t) = \frac{\nu}{\Delta x^2} \delta_{0x} v_j$, $j = 1, \dots, N$ (interior grid lines)

At $x = 0$, need (would like) a centered approximation

$$u_x(0,t) = 0 \;\rightarrow\; \frac{1}{\Delta x} \delta_{-x} v_1(t) = 0$$

recall, $\delta_{-x} v_1(t) = \frac{v_1(t) - v_0(t)}{\Delta x}$

$$\rightarrow \boxed{v_0(t) = v_1(t)}$$

$$\rightarrow \boxed{v_1'(t) = \frac{\nu}{\Delta x^2}\left(v_2 - 2v_1 + v_0\right) = \frac{\nu}{\Delta x^2}\left(v_2 - v_1\right)}$$

(notice the factor of 2 is gone)

and similarly on the other side, $\boxed{v_{N+1} = v_N}$

$$\rightarrow \boxed{v_N'(t) = \frac{\nu}{\Delta x^2}\left(v_{N-1} - v_N\right)}$$

~~Also~~

Integral Conservation for Neumann Problem

Integrate the heat equation with respect to x from $x=0$ to $x=1$.

$$\int_0^1 u_t \, dx = \nu \int_0^1 u_{xx} \, dx$$

$$\frac{d}{dt} \int_0^1 u \, dx = \underbrace{\cancel{u_x \Big]_0^1}}_{0\text{, due to Neumann BCs}}$$

$$\rightarrow \boxed{\begin{aligned} &\frac{d}{dt}\int_0^1 u \, dx = 0 \\ &\rightarrow \int_0^1 u \, dx = \text{const,} \end{aligned}}$$

<u>Question</u>: In what sense is this ~~physical~~ rule preserved in the discrete approximation?

Let us define $I(t) = \int_0^1 v(x,t)\,dx$

and define the discrete approximation (trapezoidal rule)

$$I_h(t) = \frac{\Delta x}{2} v_0(t) + \Delta x \sum_{j=1}^{N-1} v_j(t) + \frac{\Delta x}{2} v_N(t)$$

(this is for the <u>non-staggered</u> approximation)

Let us show that I_h = constant. Calculate $I_h'(t)$

$$I_h'(t) = \frac{\Delta x}{2} v_0'(t) + \Delta x \sum_{j=1}^{N-1} v_j'(t) + \frac{\Delta x}{2} v_N'(t)$$

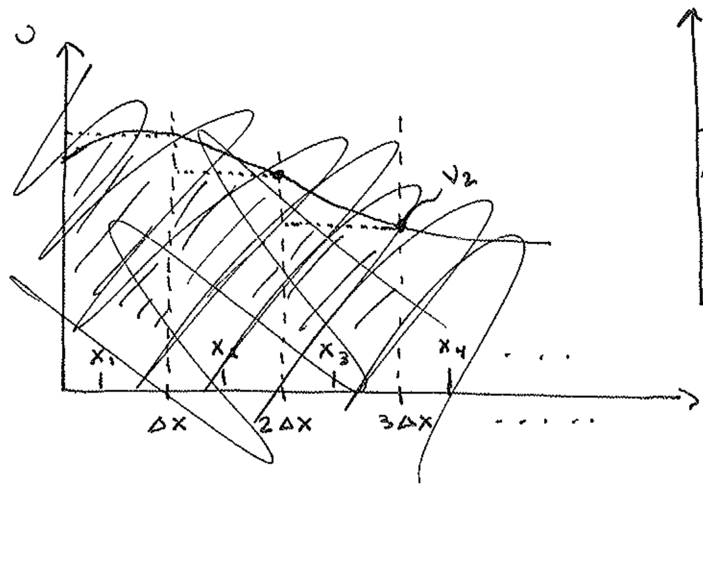
$$I_h'(t) = \frac{\Delta x}{2} \cdot \frac{\nu}{\Delta x^2}\left(2v_1 - 2v_0\right) + \Delta x \sum_{j=1}^{N-1} \frac{\nu}{\Delta x^2}\left(v_{j+1} - 2v_j + v_{j-1}\right) + \frac{\Delta x}{2}\left(2v_{N-1} - 2v_N\right)\frac{\nu}{\Delta x^2}$$

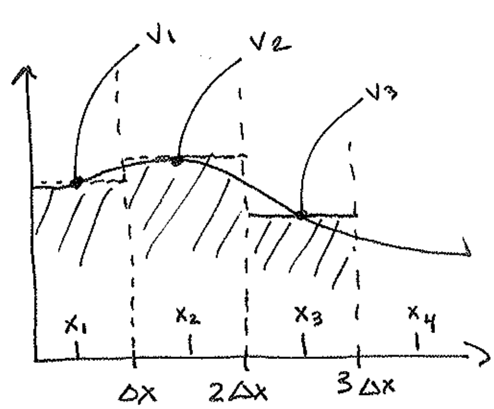
$$I_h'(t) = 0 \quad \longrightarrow \quad \boxed{I_h(t) = \text{constant}}$$

$$\longrightarrow \quad I_h(t) = I_h(0)$$

The <u>staggered</u> configuration also preserves an integral.

Define $\hat{I}_h(t) = \Delta x \sum_{j=1}^{N} v_j(t)$, Midpoint rule





Generating Discrete Approximations

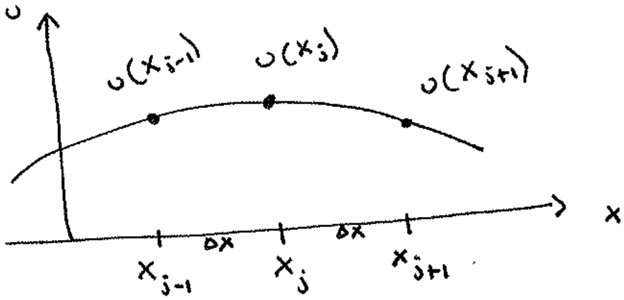
Basic problem is to find discrete approximations to derivatives that appear in PDE.

eg for the second derivative in the heat eqn, we had

$$u_{xx}(x_j, t) \simeq \frac{u(x_{j-1},t) - 2u(x_j,t) + u(x_{j+1},t)}{\Delta x^2}$$

$$x_{j-1} = x_j - \Delta x \quad , \quad x_{j+1} = x_j + \Delta x$$

Suppose we want to approximate $u_x(x_j)$



$$\text{let} \quad u_x(x_j) = \underbrace{a\,u(x_{j-1}) + b\,u(x_j) + c\,u(x_{j+1})}_{\substack{\text{Finite difference appx} \\ \text{of } u_x(x_j) \text{ for some choice} \\ \text{of coefficients } a,b,c}} + \overbrace{\tau(x_j)}^{\text{truncation error}}$$

How do you choose a, b, c?

1) we want $\tau(x_j) \to 0$ as $\max\{\Delta x_j, \Delta x_{j+1}\} \to 0$ (this is the notion of consistency)

2) we would like $\tau(x_j) \to 0$ as fast as possible, ie this is the notion of maximizing order of an approximation

3) you may want the approximation to preserve some features of the PDE

$$u_x(x_j) = a\,u(x_j - \Delta x_j) + b\,u(x_j) + c\,u(x_j + \Delta x_{j+1}) + \tau(x_j)$$

- For simplicity, set $\Delta x_j = \Delta x_{j+1} = \Delta x$
- Also note that a, b, c are independent of x_j

we can set $x_j = 0$ without loss of generality

$$\rightarrow \quad u_x(0) = a\,u(-\Delta x) + b\,u(0) + c\,u(\Delta x) + \tau$$

<u>Taylor series approach</u>

$$u_x(0) = a\left(u(0) - \Delta x\, u_x(0) + \frac{\Delta x^2}{2} u_{xx}(0) + \cdots\right) + b\,u(0) + c\left(u(0) + \Delta x\, u_x(0) + \frac{\Delta x^2}{2} u_{xx} + \cdots\right) + \tau$$

equate the coeffs of derivatives

$$\left.\begin{aligned} u(0):&\quad a + b + c = 0 \\ u_x(0):&\quad -\Delta x\, a + c\,\Delta x = 1 \\ u_{xx}(0):&\quad \frac{\Delta x^2}{2} a + \frac{\Delta x^2}{2} c = 0 \end{aligned}\right\} \text{linear system for coefficients } a, b, c$$

$$\rightarrow \quad \boxed{a = \frac{-1}{2\Delta x}, \quad b = 0, \quad c = \frac{1}{2\Delta x}}$$

The unmatched terms are absorbed into τ

$$\rightarrow \quad u_x(0) = \frac{-1}{2\Delta x}\left(u(0) - \Delta x\, u_x(0) + \frac{\Delta x^2}{2} u_{xx}(0) - \frac{\Delta x^3}{6} u_{xxx}(0) + \frac{\Delta x^4}{24} u_{xxxx}(0) + \cdots\right)$$
$$+ \frac{1}{2\Delta x}\left(u(0) + \Delta x\, u_x(0) + \frac{\Delta x^2}{2} u_{xx}(0) + \frac{\Delta x^3}{6} u_{xxx}(0) + \frac{\Delta x^4}{24} u_{xxxx}(0) + \cdots\right) + \tau$$

$$\rightarrow \quad u_x(0) = u_x(0) + \frac{\Delta x^2}{6} u_{xxx}(0) + \frac{\Delta x^4}{5!} u_{xxxxx}(0) + \cdots + \tau$$

$$\rightarrow \quad \boxed{\tau = -\frac{\Delta x^2}{3!} u_{xxx} - \frac{\Delta x^4}{5!} u_{xxxxx} + \cdots}$$

Notice, $\tau \to 0$ as $\Delta x \to 0$. The order of the approximation is 2.

As another example, consider the second derivative

$$u_{xx} = a\left(u - \Delta x u_x + \frac{\Delta x^2}{2} u_{xx} + \cdots\right) + bu + c\left(u + \Delta x u_x + \frac{\Delta x^2}{2} u_{xx} + \cdots\right) + \tau$$

match the corresponding coefficients

$$\begin{aligned} a + b + c &= 0 \\ -\Delta x a + \Delta x c &= 0 \\ \frac{\Delta x^2}{2} a + \frac{\Delta x^2}{2} c &= 1 \end{aligned} \implies \boxed{a = c = \frac{1}{\Delta x^2}, \quad b = -\frac{2}{\Delta x^2}}$$

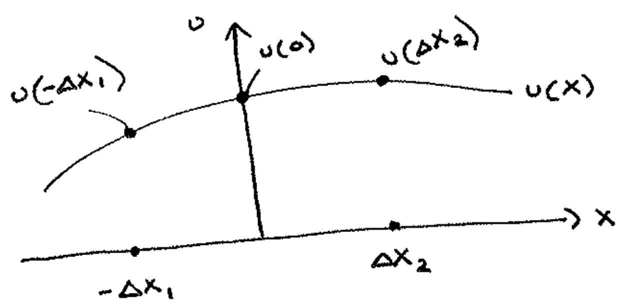
You find that $\tau = -2\left[\frac{\Delta x^2}{4!} u_{xxx} + \frac{\Delta x^4}{6!} u_{xxxxxx} + \cdots\right]$

This idea can be used to find one-sided approximations, higher order appxs (requiring more points)

9/11/03

Generating Discrete Approximations

1) Taylor series approach
2) Interpolation approach:



consider some smooth function $u(x)$, assume $u(0)$ is known. To approximate derivative at $x=0$, need local information

Note, Δx_1 and Δx_2 need not be equal. In addition, notice that the curve is relatively smooth.

Interpolate $u(x)$ using data ~~$(\Delta x_1, \Delta x_2)$~~ $(-\Delta x_1, u(-\Delta x_1))$, $(0, u(0))$, $(\Delta x_2, u(\Delta x_2))$

Call the interpolant $\tilde{u}(x)$. Suppose we want to approximate $u_x(0)$, then use $u_x(0) \simeq \tilde{u}_x(0)$

Standard approach is polynomial interpolation.

Set

$$u_x(0) = \underbrace{A\,u(-\Delta x_1) + B\,u(0) + C\,u(\Delta x_2)}_{\tilde{u}_x(0)} + E$$

$\hookrightarrow$ truncation error

This is equivalent to setting truncation error $E=0$ for $u(x) = 1, x, x^2$

$u=1$: $\quad 0 = A + B + C$

$u=x$: $\quad 1 = -A\Delta x_1 + C\Delta x_2$

$u=x^2$: $\quad 0 = A\Delta x_1^2 + C\Delta x_2^2$

Solve the linear equations to find

$$A = \frac{-1}{\Delta x_1\left(\frac{\Delta x_1}{\Delta x_2}+1\right)}, \quad C = \frac{1}{\Delta x_2\left(\frac{\Delta x_2}{\Delta x_1}+1\right)}, \quad B = -A-C$$

Notice, if $\Delta x_1 = \Delta x_2 = \Delta x$ then $A = \frac{-1}{2\Delta x}$, $C = \frac{1}{2\Delta x}$, $B = 0$

and so $u_x(0) \approx \frac{1}{2\Delta x}\Big(u(\Delta x) - u(-\Delta x)\Big)$, which is the same formula derived by Taylor series approach.

What about the error?. Need to consider higher ~~order~~ degree polynomials, $u = x^3, x^4, \ldots$ You let $E = K u^{(p)}(\xi)$ for $-\Delta x_1 \le \xi \le \Delta x_2$

Finite Volumes

Finite volume discretization is an integral form of an equation.

$$\frac{d}{dt}\int_a^b A e(x,t)\, dx = \text{rate of change of `heat' energy between } x=a,\ x=b$$

where A = constant cross-sectional area

$e(x,t)$ = `heat energy' per unit volume

Define heat flux $\Phi(x,t)$, $\left(\frac{\text{heat}}{\text{area}\cdot\text{time}}\right)$

define heat generation $f(x,t) = \left(\frac{\text{heat}}{\text{volume}\cdot\text{time}}\right)$

$$\frac{d}{dt}\int_a^b A e(x,t)\,dx = A[\Phi(a,t) - \Phi(b,t)] + \int_a^b A f(x,t)\,dx$$

A= constant, cancel out, Formula is per unit length now

$$\frac{d}{dt}\int_a^b e(x,t)\,dx = \Phi(a,t) - \Phi(b,t) + \int_a^b f(x,t)\,dx$$

Discretize this formula:

~~$a = x_0 < x_j < x$~~ choose $a = x_{j-1}$

$b = x_j$

define $\boxed{E_j(t) \equiv \frac{1}{\Delta x}\int_{x_{j-1}}^{x_j} e(x,t)\,dx, \text{ cell average energy}}$

$\boxed{F_j(t) \equiv \frac{1}{\Delta x}\int_{x_{j-1}}^{x_j} f(x,t)\,dx, \text{ cell average of heat generation term}}$

$$\rightarrow \boxed{\frac{d}{dt}E_j(t) = -\left(\frac{\Phi(x_j,t) - \Phi(x_{j-1},t)}{\Delta x}\right) + F_j(t) \quad \text{this is still an exact Formula}}$$

Define $u(x,t)$ to be a temperature

Set $e(x,t) = \underbrace{\rho(x)}_{\text{density}}\ \underbrace{c(x)}_{\text{specific heat}}\ u(x,t)$

Set $\Phi(x,t) = -\underbrace{K(x)}_{\text{heat conductivity}} u_x$

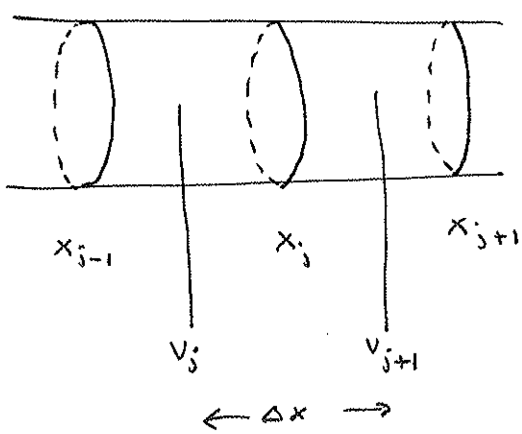
→ $E_j(t)$ is the cell average of $e(x,t)$

$$E_j(t) = \frac{1}{\Delta x}\int_{x_{j-1}}^{x_j} \rho(x)\, C(x)\, u(x,t)\, dx$$, use midpoint rule

$$E_j(t) = \rho(x_{j-1/2})\, c(x_{j-1/2})\, u(x_{j-1/2},t) + O(\Delta x^2)$$

Define $v_j(t) \simeq u(x_{j-1/2},t)$, approximation to temperature

$$\Phi(x_j,t) \simeq -K(x_j)\left(\frac{v_{j+1}(t) - v_j(t)}{\Delta x}\right)$$



The discrete equation becomes , for some range of j

Conservation Form
$$\rho(x_{j-1/2})\, c(x_{j-1/2}) \frac{d}{dt} v_j(t) = -\frac{1}{\Delta x}\left[K(x_j)\frac{v_{j+1}-v_j}{\Delta x} - K(x_{j-1})\frac{v_j - v_{j-1}}{\Delta x}\right] + F_j(t)$$

These are now a bunch of ODEs.

Convergence, Consistency, Stability

Differential equation, $Lu = F$

~~where~~
Approximate the differential equation using $L_h v_h = F_h$

where L_h is the difference operator
v_h is the grid function
F_h is the Forcing Function
h denotes the grid spacing

For example,

$$\frac{1}{\Delta t}\,\delta_{+t} v_j^n - \frac{\nu}{\Delta x^2}\,\delta_x^2 v_j^n = F(x_j, t_n)$$

The basic idea of convergence is that $v_h \to u_h$ as $h \to 0$.

Suppose we want an initial value problem. Let us not concern ourselves with boundary conditions.

$$Lu = F, \quad |x| < \infty, \quad t > 0, \quad u(x,0) = f(x)$$

The approximation is

$$L_h v_h = F_h, \quad |j| < \infty, \quad n > 0, \quad v_h = f(x_j) \text{ when } n = 0$$

with $x_j = j\Delta x$, $t_n = n\Delta t$

__Def:__ The difference approximation is said to be __convergent__ if

$$\|v_h - u_h\|_h \to 0 \quad \text{as} \quad h \to 0$$

For any $x, t \in [0, t_{Final}]$

$\|\cdot\|_h$ is some norm that is defined on the grid,

For example the infinity norm $\|e_n\|_h = \max_j |e(x_j, t)|$

The discrete approximation is <u>convergent of order (p,q)</u>

if $\quad \| v_n - u_n \| = O(\Delta x^p, \Delta t^q)$

To show that the method is convergent, often consider whether the method is consistent and stable. However, you can in some instances show convergence directly.

<u>Example</u> : <u>discretization of heat eqn.</u>

$$\frac{1}{\Delta t} \delta_{+t} v_j^n = \frac{\nu}{\Delta x^2} \delta_x^2 v_j^n$$

Show that the approximation is a convergent approximation of

$$u_t = \nu u_{xx}$$

provided that $r = \frac{\nu \Delta t}{\Delta x^2} \le \frac{1}{2}$

Set $\quad e_j^n = v_j^n - u(x_j, t_n) \qquad , \; x_j = j\Delta x , \quad t_n = n\Delta t$

denote $u(x_j, t_n) = u_j^n$ then $\quad e_j^n = v_j^n - u_j^n$

$\rightarrow \; v_j^n = u_j^n + e_j^n \quad$, substitute into Finite difference approximation

$$\frac{1}{\Delta t} \delta_{+t} (u_j^n + e_j^n) = \frac{\nu}{\Delta x^2} \delta_x^2 (u_j^n + e_j^n)$$

$$\Rightarrow \frac{1}{\Delta t} \delta_{+t} e_j^n - \frac{\nu}{\Delta x^2} \delta_x^2 e_j^n = - \left[\frac{1}{\Delta t} \delta_{+t} u_j^n - \frac{\nu}{\Delta x^2} \delta_x^2 u_j^n \right]$$

We have shown (9/4/03) that

$$\frac{1}{\Delta t} \delta_{+t} u_j^n - \frac{\nu}{\Delta x^2} \delta_x^2 u_j^n = O(\Delta t, \Delta x^2)$$

$$\Rightarrow \quad \delta_{+t} e_j^n - r \delta_x^2 e_j^n = O(\Delta t^2, \Delta t \Delta x^2)$$

$$\rightarrow \quad e_j^{n+1} = e_j^n + r \delta_x^2 e_j^n + O(\Delta t^2, \Delta t \Delta x^2)$$

Take the absolute value, use Δ inequality

$$e_j^{n+1} = e_j^n + r(e_{j-1}^n - 2re_j^n + e_{j+1}^n) + O(\Delta t^2, \Delta t\,\cancel{\Delta x^2})$$

$$\rightarrow |e_j^{n+1}| \le |1-2r||e_j^n| + r|e_{j-1}^n| + r|e_{j+1}^n| + A(\Delta t^2, \Delta t\,\cancel{\Delta x^2})$$

if $r \le \frac{1}{2}$, $|1-2r| = 1-2r$

define $E^n = \max_j |e_j^n|$

$$|e_j^{n+1}| \le (1-2r)E^n + rE^n + rE^n + A(\Delta t^2, \Delta t\,\cancel{\Delta x^2})$$

$\rightarrow |e_j^{n+1}| \le E^n + A(\Delta t^2, \Delta t\Delta x^2)$ for all j and in particular for that value of j where e is max

$$\begin{aligned}
\rightarrow E^{n+1} &\le E^n + A(\Delta t^2, \Delta t\Delta x^2) \\
&\le E^{n-1} + 2A(\Delta t^2, \Delta t\Delta x^2) \\
&\le E^{n-2} + 3A(\Delta t^2, \Delta t\Delta x^2) \\
&\vdots \\
&\le E^0 + (n+1)A(\Delta t^2, \Delta t\Delta x^2)
\end{aligned}$$

apply recursively

but $E^0 = 0$, therefore $E^n \le nA(\Delta t^2, \Delta t\Delta x^2)$

but $n\Delta t \le t_{Final}$ $\rightarrow$ $\boxed{E^n \le K(\Delta t, \Delta x^2) \rightarrow \text{convergent}}$

9/15/03

We're discussing convergence, consistency, stability

PDE, initial value problem, $Lu = F$, $|x| < \infty$, $t > 0$

initial condition, $u(x,0) = f(x)$

we have a discrete approximation, $L_h v_h = F_h$

for some grid (x_j, t_n), $x_j = j\Delta x$, $t_n = n\Delta t$, $|j| < \infty$, $n > 0$

$v_h = f(x_j)$ at $n = 0$

The idea of convergence is that the numerical approximation approaches the exact solution restricted to the grid as $h \to 0$

$$\boxed{v_h \to u_h \text{ as } h \to 0}$$

that was the case for a pure IVP, but if you have a pure BVP, the idea is essentially the same with a few 'rinkles.'

- Consider the BVP,

$$\boxed{\begin{array}{l} Lu = F, \quad \cancel{0 \le x \le 1}\ 0 < x < 1, \ t > 0 \\ u(x,0) = f(x) \ + \text{BCs at } x = 0 \text{ and } x = 1 \end{array}}$$

Absorb the BCs into the problem.

$$\rightarrow \boxed{\begin{array}{l} \hat{L}u = \hat{F}, \quad 0 \le x \le 1, \ t > 0 \quad \text{(notice problem now includes boundary)} \\ u(x,0) = f(x) \end{array}}$$

Example:
$$u_t = \nu u_{xx} + F(x,t)$$
$$u(0,t) = a(t)$$
$$u_x(1,t) - u(1,t) = b(t)$$

Now define $\hat{L}$ and $\hat{F}$

$$\hat{L}u = \begin{cases} u_t - \nu u_{xx}, & 0 < x < 1 \\ u, & x = 0 \\ u_x - u, & x = 1 \end{cases}, \quad \hat{F} = \begin{cases} F, & 0 < x < 1 \\ a, & x = 0 \\ b, & x = 1 \end{cases}$$

discrete approximation: $\hat{L}_h v_h = \hat{F}_h$, $0 \le j \le N$, $n > 0$

$$v_h = f(x_j) \quad , n = 0$$

to define convergence, introduce a sequence $\{\Delta x_k\}_{k=1}^{\infty}$ such that $\Delta x_k \to 0$ as $k \to \infty$. This is required due to finite interval.

Therefore we need $\boxed{v_{h_k} \to u_{h_k} \text{ as } k \to \infty}$

So the idea between pure IVPs and BVPs (concerning convergence) is the same!

Consistency

The idea for consistency is that the discrete approximation should approach the PDE (+BCs) as $h \to 0$.

Def A discrete approximation $L_h v_h = F_h$ of a differential equation $Lu = F$ is consistent if

$$\| (L\phi - F)_h - (L_h \phi_h - F_h) \|_h \to 0$$

as $h \to 0$ for any smooth function ϕ.

Note, ϕ is not necessarily a solution of PDE. However, often take $\phi = u$, the solution of the PDE. then $L\phi - F = 0$ and then

$$L_h \phi_h - F_h = L_h u_h - F_h = -\tau_h$$, the local truncation error

Therefore this definition of consistency really amounts to whether local truncation error vanishes as grid is refined $\|\tau_h\| \to 0$ as $h \to 0$

→ The amount by which the exact solution fails to satisfy the difference approximation must tend to zero as $h \to 0$.

The <u>order of accuracy</u> of an approximation depends on the rate at which $\tau_h \to 0$.

eg if $\|\tau_h\|_h = O(\Delta x^p, \Delta x^q)$

then the order of accuracy is p^{th} order in space and q^{th} order in time.

<u>Example</u> consider the discretization

$$\boxed{\frac{1}{\Delta t}\delta_{+t} v_j^n = \frac{\nu}{2\Delta x^2}\delta_x^2\left(v_j^n + v_j^{n+1}\right) \quad \text{Crank-Nicholson method}}$$

show that the discrete approximation is consistent with the equation $u_t = \nu u_{xx}$ and determine the order of accuracy.

<u>Define truncation error,</u> τ_j^n

$$\tau_j^n = \frac{1}{\Delta t}\delta_{+t}u(x_j,t_n) - \frac{\nu}{2\Delta x^2}\delta_x^2\Big(u(x_j,t_n) + u(x_j,t_{n+1})\Big)$$

<u>First term:</u>

$$\delta_{+t}u_j^n = u_j^{n+1} - u_j^n = \Delta t\, u_t + \frac{\Delta t^2}{2}u_{tt} + \frac{\Delta t^3}{6}u_{ttt} + \cdots$$

$$\rightarrow \quad \frac{1}{\Delta t}\delta_{+t}u_j^n = u_t + \frac{\Delta t}{2}u_{tt} + \frac{\Delta t^2}{6}u_{ttt} + \cdots$$

<u>Second term:</u>

$$\frac{1}{\Delta x^2}\delta_x^2 u_j^n = \frac{1}{\Delta x^2}\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right)$$

$$= \frac{1}{\Delta x^2}\left(\Delta x^2 u_{xx} + \frac{\Delta x^4}{12}u_{xxxx} + O(\Delta x^6)\right)$$

$$= u_{xx} + \frac{\Delta x^2}{12}u_{xxxx} + O(\Delta x^4)$$

Third term:

$$\frac{1}{\Delta x^2}\delta_x^2 u_j^{n+1} = \left[u_{xx} + \frac{\Delta x^2}{12}u_{xxxx} + O(\Delta x^4)\right]_{(x_j,t_{n+1})}$$

$$= u_{xx} + \Delta t\, u_{xxt} + \frac{\Delta t^2}{2}u_{xxtt} + O(\Delta t^3)$$

$$+ \frac{\Delta x^2}{12}u_{xxxx} + \frac{\Delta x^2\Delta t}{12}u_{xxxxt} + \cancel{O(\Delta t^3)} + O(\Delta x^2, \Delta t^2)$$

$$+ O(\Delta x^4)$$

$$\rightarrow \tau_j^n = \cancel{u_t} + \frac{\Delta t}{2}u_{tt} + \frac{\Delta t^2}{6}u_{ttt} - \frac{\nu}{2}\left[\cancel{2u_{xx}} + \Delta t\, u_{xxt} + \frac{2\Delta x^2}{12}u_{xxxx} + O(\Delta t^2, \Delta x^2\Delta t, \Delta x^4)\right]$$

$$u_{tt} = \nu u_{xxt}$$

$\rightarrow \tau_j^n = O(\Delta t^2, \Delta x^2)$ → therefore consistent because as $\Delta x \to 0$, $\Delta t \to 0$, $\tau_j^n \to 0$ and 2nd order in both time and space

Stability

Stability is a form of well-posedness for difference equations.

Initial-value problems: $Lu = F$, $|x| < \infty$, $t > 0$, $u(x,0) = f(x)$

discrete approximation: $L_h v_h = F_h$, $|j| < \infty$, $n > 0$, $v_h = f(x_j)$ at $n = 0$

Suppose the problems are linear. Define an error e_h,

$$e_h = v_h - u_h$$

Error solves $L_h e_h = F_h - L_h u_h = \tau_h$, with $e_h = 0$ at $n = 0$

If the discrete approximation is consistent then $\tau_h \to 0$ as $h \to 0$.

we would like that $e_h \to 0$ as $h \to 0$, which would imply convergence.

For stability, we need only consider the homogeneous problem, $L_h v_h = 0$, v_h given at $n=0$.

Include a superscript n to denote time:

$$L_h v_h^n = 0 \quad , \quad \text{with } v_h^0 \text{ given}$$

We require this equation to be well-behaved for stability.

> **Def** A difference scheme $L_h v_h^n = 0$ is stable if $K \geq 0$, $\beta \geq 0$ exist such that
> $$\| v_h^n \|_h \leq \| v_h^0 \|_h K e^{\beta t} \quad , \quad t = n \Delta t$$

Suppose you're only interested in integrating to some finite time, ie you restrict $t \in (0, t_{final})$ then

$$\| v_h \|_h \leq \hat{K} \| v_h^0 \|_h$$

or if the difference approximation is meant to be an approximation of a differential equation whose solutions are "bounded", "stable", then you might apply a stability criterion such as

$$\| v_h^n \|_h \leq K \| v_h^0 \| \text{ for all } t$$

The bottom line is that there a various definitions of stability that make sense depending on the situation.

Often, time-stepping schemes are two level schemes of the form $v_h^{n+1} = Q_h v_h^n$, where Q_h is some difference operator. The general form for stability for the two level schemes are

$$\| Q_h^n \| \leq K e^{\beta t} \quad , \quad t = n \Delta t \quad \text{or} \quad \| Q_h \| \leq 1 + \alpha \Delta t \, , \; \alpha = \text{const.}$$

$$\| Q_h \| = \max \frac{\| Q_h v_h \|}{\| v_h \|} \text{ for all } v_h$$

Reduce the question of stability to a consideration of the behavior of the stepping function.

Example: consider the scheme $v_j^{n+1} = v_j^n - \sigma \delta_{-x} v_j^n$

Show that this scheme is stable if $0 \le \sigma \le 1$

Take absolute value of both sides:

$$\cancel{|v_j^{n+1}| \le |v_j^n| + |\sigma| \, |\delta_{-x} v_j^n|}$$

$$\cancel{\le |v_j^n| + |\sigma| \left(|v_j^n| + |v_{j-1}^n| \right)}$$

$$v_j^{n+1} = v_j^n - \sigma \left(v_j^n - v_{j-1}^n \right)$$

$$v_j^{n+1} = (1-\sigma) v_j^n + \sigma v_{j-1}^n$$

$$\rightarrow |v_j^{n+1}| \le (1-\sigma) |v_j^n| + \sigma |v_{j-1}^n|$$

define $Z^n = \max_j |v_j^n|$

$$|v_j^{n+1}| \le (1-\sigma) Z^n + \sigma Z^n = Z^n, \quad \forall j$$

For j such that $|v_j^{n+1}| = Z^{n+1}$

$$\rightarrow Z^{n+1} \le Z^n \le \cdots \le Z^0 \rightarrow \text{stability}$$

9/18/03

Lax Theorem (Two versions)

Lax Equivalence Theorem

A consistent, two-level difference scheme for a well posed linear IVP is convergent if and only if it is stable.

Lax Theorem

If a two-level difference scheme having the form: $v_h^{n+1} = Q_h v_h^n + \Delta t\, G_h^n$, $n\Delta t = t \le t_{Final}$, is accurate of order (p, q) in the norm $\|\cdot\|_h$ to a well-posed IVP and is stable with respect to $\|\cdot\|_h$, then it is convergent of order (p,q) with respect to $\|\cdot\|_h$. (Q is a linear operator)

consistency + stability = convergence

why does this work?

let $w_h^n = v_h^n - u_h^n$ = error

$$\rightarrow\ w_h^{n+1} = Q_h w_h^n + \Delta t\, G_h^n - u_h^{n+1}$$

$$w_h^{n+1} = Q_h(w_h^n + u_h^n) + \Delta t\, G_h^n - u_h^{n+1}$$

$$w_h^{n+1} = Q_h w_h^n + \underbrace{\left[\Delta t\, G_h^n + Q_h u_h^n - u_h^{n+1}\right]}_{\Delta t\, \tau_h^n \text{, truncation error}}$$

$$w_h^{n+1} = Q_h w_h^n + \Delta t\, \tau_h^n$$

by recursion,

$$w_h^n = Q_h w_h^{n-1} + \Delta t\, \tau_h^{n-1} = Q_h\left(Q_h w_h^{n-2} + \Delta t\, \tau_h^{n-2}\right) + \Delta t\, \tau_h^{n-1}$$

$$= \dots = |Q_h|^n w_h^0 + \Delta t \sum_{k=1}^{n} Q_h^{k-1} \tau_h^{n-k}$$

$$\rightarrow\ \boxed{w_h^n = \Delta t \sum_{k=1}^{n} Q_h^{k-1} \tau_h^{n-k}}$$

Take $\|\cdot\|_h$ and use

$$\|\tau_h^n\|_h \le A(\Delta t^q + \Delta x^p) \quad \forall n \quad \text{(From consistency)}$$

~~From consistency~~

From stability, we have $\|Q_h^n\|_h \le K e^{\beta t}$

$$\rightarrow \|w_h^n\| \le \Delta t\, n\, K e^{\beta t} A(\Delta x^p + \Delta t^q)$$

$$\rightarrow \boxed{\|w_h^n\| \le t_{Final} K e^{\beta t} A(\Delta x^p + \Delta t^q)}$$

Fourier Stability Analysis

Let us consider an IVP for the heat equation

$u_t = \nu u_{xx}$, $|x| < \infty$, $t > 0$, $u(x,0) = f(x)$

$$\mathcal{F}[u(x,t)] = \hat{u}(\omega,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\omega x} u(x,t)\, dx$$

$$\mathcal{F}[u_t(x,t)] = \hat{u}_t(\omega,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\omega x} u_t(x,t)\, dx$$

$$\mathcal{F}[u_{xx}(x,t)] = (i\omega)^2 \hat{u}(\omega,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\omega x} u_{xx}(x,t)\, dx$$

the transformed PDE becomes $\hat{u}_t = -\nu^2 \omega^2 \hat{u}$

with initial condition $\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$

inverse transform, $\mathcal{F}^{-1}[\hat{u}(\omega,t)] \doteq u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega x} \hat{u}(\omega,t)\, d\omega$

Solving for $u(x,t)$, we obtain

$$u(x,t) = \frac{1}{\sqrt{4\pi \nu t}} \int_{-\infty}^{\infty} f(x_0)\, e^{-\frac{(x-x_0)^2}{4\nu t}}\, dx_0$$

For constant coefficient linear difference equations a similar analysis involving discrete Fourier transforms is used to analyze stability.

Define the discrete Fourier transform:

$$F v_h = \hat{v}(\xi) = \frac{1}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} e^{-ij\xi}\, v_j \quad , \quad \xi \in [-\pi, \pi]$$

The inverse Fourier transform

$$F^{-1}\hat{v}(\xi) = v_j = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{ij\xi}\, \hat{v}(\xi)\, d\xi$$

The norm is preserved under the transformations

$$\| v_h \|^2 = \sum_{j=-\infty}^{\infty} |v_j|^2 = \int_{-\pi}^{\pi} |\hat{v}(\xi)|^2\, d\xi = \| \hat{v}(\xi) \|^2$$

$$\rightarrow \boxed{\| v_h \| = \| \hat{v}(\xi) \|} \qquad \| v_h^n \| \le K e^{\beta t} \| v_h^0 \| , \quad \| \hat{v}^n \| \le K e^{\beta t} \| \hat{v}^0 \|$$

Example:

Consider the stability of the difference scheme

$$v_j^{n+1} = (1-2r)\, v_j^n + r\left(v_{j-1}^n + v_{j+1}^n\right), \quad r = \frac{\nu \Delta t}{\Delta x^2}$$

Calculate some DFTs:

$$F\, v_j^{n+1} = \frac{1}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} e^{-ij\xi} v_j^{n+1} = \hat{v}^{n+1}(\xi)$$

$$F\, v_j^{n} = \frac{1}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} e^{-ij\xi} v_j^{n} = \hat{v}^{n}(\xi)$$

$$F\, v_{j-1}^{n} = \frac{1}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} e^{-ij\xi} v_{j-1}^{n} = \frac{1}{\sqrt{2\pi}} \sum_{m=-\infty}^{+\infty} e^{-i(m+1)\xi} v_m^n$$

$$= \frac{e^{-i\xi}}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} e^{-ij\xi} v_j^n = e^{-i\xi}\, \hat{v}^n(\xi)$$

and similarly, $F\, v_{j+1}^n = e^{i\xi}\, \hat{v}^n(\xi)$

substitute into difference scheme

$$\hat{v}^{n+1}(\xi) = \left[(1-2r) + r\left(e^{-i\xi} + e^{i\xi}\right)\right] \hat{v}^n(\xi)$$

$$\hat{v}^{n+1}(\xi) = \left[1-2r + 2r\cos\xi\right] \hat{v}^n(\xi) = \left(1-2r(1-\cos\xi)\right)\hat{v}^n(\xi)$$

$\Rightarrow$ $\hat{v}^{n+1}(\xi) = \rho(\xi)\,\hat{v}^n(\xi)$ where $\rho(\xi)$ is the "symbol" of the difference equation

$\rightarrow$ $\rho(\xi) = 1-2r(1-\cos\xi) = 1-4r\sin^2\left(\frac{\xi}{2}\right)$

If we require $|\rho(\xi)| \le 1$ for $\xi \in [-\pi, \pi]$ then we have stability (with $K=1$, $\beta=0$)

Note, for $\xi \in [-\pi,\pi] \rightarrow \sin^2\left(\frac{\xi}{2}\right) \in [0,1] \rightarrow |\rho(\xi)| \le 1 \Rightarrow 1-4r \ge -1$

$\rightarrow$ $\boxed{r \le \frac{1}{2} \text{, as before}}$

we had $\hat{v}^{n+1}(\xi) = \rho(\xi)\,\hat{v}^n(\xi)$, $\hat{v}^n(\xi) = [\rho(\xi)]^n v^0(\xi)$

$\rightarrow \|\hat{v}^n(\xi)\| = \|\rho(\xi)^n v^0(\xi)\| \le \|\rho(\xi)^n\|\,\|v^0(\xi)\|$

but $\|\rho(\xi)^n\| \le K e^{\beta t}$, $t = n\Delta t$

For stability, clearly if $|\rho(\xi)| \le 1$, the inequality for stability is satisfied.

__Example__ : __convection-diffusion__

Analyze the stability of $v_j^{n+1} = v_j^n - \frac{\sigma}{2}\delta_0 v_j^n + r\,\delta_x^2 v_j^n$

which approximates $u_t + c u_x = \nu u_{xx}$, $\sigma = \frac{c\Delta t}{\Delta x}$, $r = \frac{\nu \Delta t}{\Delta x^2}$

Take DFT :

$$\hat{v}^{n+1} = \hat{v}^n - \frac{\sigma}{2}\left(e^{i\xi} - e^{-i\xi}\right)\hat{v}^n + r\left(e^{i\xi} - 2 + e^{-i\xi}\right)\hat{v}^n$$

$$\hat{v}^{n+1} = \underbrace{\left[1 + i\sigma\sin(\xi) + 2r(\cos\xi - 1)\right]}_{\rho(\xi)}\hat{v}^n$$

Notice $\rho(\xi)$ is complex

$$|\rho(\xi)|^2 = \left(1 - 2r(1-\cos\xi)\right)^2 + \sigma^2\sin^2\xi$$

Let $z = \cos\xi \rightarrow |\rho(\xi)|^2 = (1 - 2r(1-z))^2 + \sigma^2(1-z^2)$

$$|\rho(\xi)| \le 1 \rightarrow \cancel{1} - 4r(1-z) + 4r^2(1-z)^2 + \sigma^2(1-z^2) \le \cancel{1}$$

$$\rightarrow -\underbrace{4r + 4r^2(1-z) + \sigma^2(1+z)}_{g(z)} \le 0$$

since $g(z)$ is linear, need only check endpoints

$g(1) = -4r + 2\sigma^2 \le 0 \rightarrow \sigma^2 \le 2r$

$g(-1) = -4r + 8r^2 \le 0 \rightarrow 2r \le 1$

$\rightarrow \sigma^2 \le 2r \le 1$ for stability

Lax Equivalence Theorem

- A consistent, two-level difference scheme for a well-posed linear IVP is convergent if and only if it is stable.

Lax Theorem (special case of above theorem)

If a two-level difference scheme having the form $v_h^{n+1} = Q_h v_h^n + \Delta t\, G_h^n$, $n\Delta t \le t_{Final}$ is accurate of order (p,q) in the grid norm $\|\cdot\|_h$ to a well posed linear IVP and is stable wrt $\|\cdot\|_h$, then it is convergent of order (p,q).

$\Rightarrow$ $\boxed{\text{consistency} + \text{stability} \Rightarrow \text{convergent}}$

Fourier Stability Analysis

discrete Fourier transform (DFT):

$$\hat{v}(\xi) = \frac{1}{\sqrt{2\pi}} \sum_{j=-\infty}^{\infty} e^{-ij\xi} v_j \, , \quad \xi \in [-\pi, \pi]$$

$\boxed{\hat{v}(\xi) = F v_j}$ - discrete FT of grid function v_j

$$v_j = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \hat{v}(\xi) e^{ij\xi} \, d\xi \quad \text{for all } j$$

$\boxed{v_j = F^{-1} \hat{v}}$

where we assume v_j decays sufficiently quickly in order for the summation to exist

Parseval's Equality.

$$\|v_h\|^2 = \sum_{j=-\infty}^{\infty} |v_j|^2 = \int_{-\pi}^{\pi} |\hat{v}(\xi)|^2 \, d\xi = \|\hat{v}(\xi)\|^2$$

$\rightarrow$ if we can show regularity of ~~own~~ $\hat{v}$ with respect to $\|\cdot\|^2 \Rightarrow$ regularity of v_h in the analagous norm.

Example An implicit scheme:

$$v_j^{n+1} = v_j^n + r\,\delta_x^2 v_j^{n+1} \quad , \quad r > 0$$

this is a discretization of heat equation with $r = \frac{\Delta t}{\Delta x^2}\nu$.

Take a DFT of the equation:

$$F v_j^{n+1} = F v_j^n + r F \delta_x^2 v_j^{n+1}$$

$$\hat{v}^{n+1} = \hat{v}^n + rF\left(v_{j-1}^{n+1} - 2v_j^{n+1} + v_{j+1}^{n+1}\right)$$

$$\hat{v}^{n+1} = \hat{v}^n + rF v_{j-1}^{n+1} - 2rF v_j^{n+1} + rF v_{j+1}^{n+1}$$

$$\hat{v}^{n+1} = \hat{v}^n + r e^{-i\xi} \hat{v}^{n+1} - 2r\hat{v}^{n+1} + r e^{i\xi} \hat{v}^{n+1}$$

$$\hat{v}^{n+1} = \hat{v}^n + r\left(e^{i\xi} + e^{-i\xi} - 2\right)\hat{v}^{n+1}$$

$$\hat{v}^{n+1} = \hat{v}^n + r\left(2\cos\xi - 2\right)\hat{v}^{n+1}$$

$$\left[1 + 2r(1 - \cos\xi)\right]\hat{v}^{n+1} = \hat{v}^n$$

$$\rightarrow \quad \hat{v}^{n+1} = \frac{\hat{v}^n}{1 + 2r(1-\cos\xi)}$$

$$\rho(\xi) = \frac{1}{1 + 2r(1-\cos\xi)}$$

, the "symbol" of the difference scheme, describes growth or decay of discrete grid function in transformed space

→ notice, $1 \le 1 + 2r(1-\cos\xi) \le 1+4r$

→ $|\rho(\xi)| \le 1$

→ $\|\hat{v}^n\| \le \|\hat{v}^0\|$ → stability, independent of r

↓

$\boxed{\text{unconditionally stable}}$

Implicit methods often have favorable stability properties as compared to explicit schemes.

Example: Exponential growth or decay. Consider the heat equation with a linear source term.

$$u_t = \nu u_{xx} + bu$$

(growth or decay depending on sign of b)

Approximate using:

$$v_j^{n+1} = v_j^n + r\,\delta_x^2 v_j^{n+1} + b\Delta t\, v_j^n$$

Analyze stability: take DFT

$$\hat{v}^{n+1} = \hat{v}^n - 2r(1-\cos\xi)\hat{v}^{n+1} + b\Delta t\,\hat{v}^n$$

$$\rightarrow \hat{v}^{n+1} = \underbrace{\frac{1+b\Delta t}{1+2r(1-\cos\xi)}}_{\rho(\xi)}\hat{v}^n$$

$|\rho(\xi)| \le |1+b\Delta t|$

$-2x < 1$
$x < -\frac{1}{2}$

there are two cases:

if $b>0$, (growth) → $|\rho| \le 1+b\Delta t$ → stable (because stability allows some growth)

if $b<0$, (decay) → $|\rho| \le |1-(-b)\Delta t|$

→ $-b\Delta t \le 2$ → $\Delta t \le \frac{2}{-b}$ $(b<0)$

von Neumann condition

Using a Fourier analysis, you obtain

$$\hat{v}^{n+1} = \rho(\xi)\,\hat{v}^{n}$$

The von Neumann stability condition is that

$$|\rho(\xi)| \le 1 + c\,\Delta t \quad , \text{ for } \xi \in [-\pi, \pi]$$

where c is a constant independent of Δx and Δt and for $\Delta x, \Delta t$ sufficiently small.

Tighter condition: $|\rho(\xi)| \le 1$ (for no growth)

Fourier Mode Analysis

inverse DFT: $v_j^n = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \hat{v}^n(\xi)\, e^{i\xi j}\, d\xi \quad \forall j$

the integral represents a sum of Fourier modes of the form

$$\underbrace{\left(\frac{1}{\sqrt{2\pi}}\,\hat{v}^n(\xi)\right)}_{\text{temporal component}}\underbrace{\left(e^{i\xi j}\right)}_{\text{spatial component}}$$

spatial component: $e^{i\xi j} = e^{ikx_j}$, $x_j = j\Delta x$, $k = \xi/\Delta x$

$$= \cos(kx_j) + i\sin(kx_j)$$

these represent spatial oscillations where k is a spatial frequency = wave number.

Recall, $\xi \in [-\pi, \pi] \rightarrow k \in \left[\frac{-\pi}{\Delta x}, \frac{\pi}{\Delta x}\right] \rightarrow$ Finite range of wave numbers, (higher values of k are aliased to lower ones) the temporal component: $\frac{1}{\sqrt{2\pi}} \hat{v}^n(\xi) = \frac{1}{\sqrt{2\pi}} \underbrace{\left[a(k\Delta x)\right]}^{n} \hat{v}^0(\xi)$ ← power, not index (25)

"growth" factor
or
"amplitude" factor
tells you how the mode
with wave number k grows
or decays ξ

<u>The idea</u>: in a Fourier mode analysis, consider solutions of the difference equation of the form

$$\underbrace{v_j^n = a^n e^{ikx_j}}_{\text{separable solution of difference eqn.}}, \quad x_j = j\Delta x,\ k = \text{wave number}$$
a = amplitude factor (complex)

von Neumann condition in this analysis becomes
$|a| \le 1 + c\Delta t$, for $|k\Delta x| \le \pi$

or more stringently, $|a| \le 1$, for $|k\Delta x| \le \pi$

(the dimensions of k is $1/\text{Length}$, such that $k\Delta x$ is non-dimensional)

$k\Delta x$ - "gridnumber"

<u>Example</u>: Convection - Diffusion problem

$$v_j^{n+1} = v_j^n - \frac{\sigma}{2}\delta_{0x} v_j^n + r\,\delta_x^2 v_j^n$$

explicit scheme for $u_t + cu_x = \nu u_{xx}$.

$\sigma = \frac{\Delta t c}{\Delta x}$ (positive or negative), $r = \frac{\Delta t \nu}{\Delta x^2} > 0$

Notice: if diffusive term is not present, this scheme is unstable

$$v_j^{n+1} = v_j^n - \frac{\sigma}{2}\,\delta_{0x} v_j^n + r\,\delta_x^2 v_j^n$$

Example, continued

<u>Mode analysis</u>, $v_j^n = a^n e^{ikx_j}$

$$\rightarrow \quad v_j^{n+1} = a^{n+1} e^{ikx_j} = a\, v_j^n$$

$$\rightarrow \quad \delta_{0x} v_j^n = v_{j+1}^n - v_{j-1}^n = a^n e^{ikx_{j+1}} - a^n e^{ikx_{j-1}}$$

$$= a^n e^{ik(x_j+\Delta x)} - a^n e^{ik(x_j-\Delta x)}$$

$$= \cancel{e^{ikx_j}}\, e^{ik\Delta x} v_j^n - e^{-ik\Delta x} v_j^n$$

$$= 2i \sin(k\Delta x)\, v_j^n$$

$$\delta_x^2 v_j^n = v_{j+1}^n - 2v_j^n + v_{j-1}^n = e^{ik\Delta x} v_j^n - 2v_j^n + e^{-ik\Delta x} v_j^n$$

$$= -2\left(1-\cos(k\Delta x)\right) v_j^n$$

$$\rightarrow \quad a v_j^n = v_j^n - \frac{\sigma}{2}\left(2i\sin(k\Delta x)\, v_j^n\right) - 2r\left(1-\cos(k\Delta x)\right) v_j^n$$

$$\boxed{a = 1 - i\sigma\sin(k\Delta x) - 2r\left(1-\cos(k\Delta x)\right)}$$

For no growth we would like $|a| \leq 1 \quad \forall\ |k\Delta x| \leq \pi$

$$|a|^2 = \left(\mathrm{Re}(a)\right)^2 + \left(\mathrm{Im}(a)\right)^2 \leq 1$$

trick is to let $\nu = \cos(k\Delta x)$ and $\sin(k\Delta x) = \sqrt{1-\nu^2}$ and so on...

Stability Analysis for an Initial-BVP problem

Example: $u_t = \nu u_{xx}$, $0 \le x \le 1$, $t \ge 0$

$u(x,0) = f(x)$

$u(0,t) = a(t)$ } Dirichlet BCs

$u(1,t) = b(t)$

discrete approximation:

$$\frac{1}{\Delta t}\delta_{+t} v_j^n = \frac{\nu}{\Delta x^2}\delta_x^2 v_j^n \quad - \quad 1 \le j \le N-1, \ n \ge 0$$

$x_j = j\Delta x$, $\Delta x = \frac{1}{N}$, $t_n = n\Delta t$

$v_j^0 = f(x_j)$, $0 \le j \le N$

$v_0^n = a(t_n)$, $v_N^n = b(t_n)$, $n > 0$

Midterm Exam: Tentative Date, tues 10/14/03

what is the stability of the approximation? Consider a perturbation to v_j^n, it would solve the discrete problem with homogeneous boundary conditions.

Call the perturbed solution $\tilde{v}_j^n$. It solves

$$\tilde{v}_j^{n+1} = \tilde{v}_j^n + r\,\delta_x^2 \tilde{v}_j^n , \quad r = \frac{\nu\Delta t}{\Delta x^2}$$

with $\tilde{v}_j^0$ given and $\tilde{v}_0^n = \tilde{v}_N^n = 0$.

Consider the behavior of this perturbation:

$$\tilde{V}_1^{n+1} = \tilde{V}_1^n + r\left(\underbrace{\cancel{\tilde{V}_0^n}}_{0,\text{ due to BC}} - 2\tilde{V}_1^n + \tilde{V}_2^n\right)$$

$$\tilde{V}_1^{n+1} = (1-2r)\tilde{V}_1^n + r\tilde{V}_2^n$$

$$\tilde{V}_2^{n+1} = (1-2r)\tilde{V}_2^n + r\tilde{V}_1^n + r\tilde{V}_3^n$$

$$\vdots$$

$$\tilde{V}_{N-1}^{n+1} = (1-2r)\tilde{V}_{N-1}^n + r\tilde{V}_{N-2}^n + \underbrace{\cancel{r\tilde{V}_N^n}}_{0}$$

this forms a linear system for $\underline{\tilde{V}}^{n+1} = \begin{bmatrix} \tilde{V}_1^{n+1} \\ \tilde{V}_2^{n+1} \\ \vdots \\ \tilde{V}_{N-1}^{n+1} \end{bmatrix}$

$$\Rightarrow \quad \underline{\tilde{V}}^{n+1} = \underline{\underline{A}}\,\underline{\tilde{V}}^n$$

where
$$A = \begin{bmatrix} 1-2r & r & & & \\ r & 1-2r & r & & \\ & r & 1-2r & \ddots & \\ & & \ddots & \ddots & \ddots \\ & & r & 1-2r & r \\ & & & r & 1-2r \end{bmatrix}$$

TRIDIAGONAL MATRIX

$\Rightarrow$ repeated application shows that

$$\underbrace{\underline{\tilde{V}}^{n}}_{\text{index}} = \underbrace{\underline{\underline{A}}^{n}}_{\text{power}}\,\underbrace{\underline{\tilde{V}}^{0}}_{\text{index}}$$

Therefore for the perturbation to be "well behaved," we want

$$\| A^n \| \leq 1$$

The task is to bound powers of $\underline{\underline{A}}$, the growth matrix, independent of N (spatial grid) and Δt. May be difficult to do.

Notice that $\underline{\underline{A}}$ is symmetric, for our example.

$\Rightarrow$ a orthogonal matrix $\underline{\underline{R}}$ exists such that $\underline{\underline{R}}^T \underline{\underline{A}}\, \underline{\underline{R}} = \underline{\underline{\Lambda}}$ where $\underline{\underline{\Lambda}}$ is a matrix whose diagonal entries are real and are the eigenvalues of $\underline{\underline{A}}$

$$\underline{\underline{\Lambda}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}, \quad \underline{\underline{R}} = \begin{bmatrix} | & | & & | \\ \underline{r}_1 & \underline{r}_2 & \cdots & \underline{r}_n \\ | & | & & | \end{bmatrix}$$

λ_j - eigenvectors of $\underline{\underline{A}}$, $\underline{r}_j$ - right eigenvectors of $\underline{\underline{A}}$

For this case, the <u>2-norm</u> of $\underline{\underline{A}}$ is equal to:

$$\|\underline{\underline{A}}\| = \|\underline{\underline{R}}\,\underline{\underline{\Lambda}}\,\underline{\underline{R}}^T\| = \|\underline{\underline{\Lambda}}\| = \max_j |\lambda_j| \leftarrow \text{spectral radius}$$

The problem of stability becomes analyzing the eigenvalues of the growth matrix. This is true for all symmetric matrices $\underline{\underline{A}}$.

Let us calculate eigenvalues for $\underline{\underline{A}}$: $\underline{\underline{A}}\,\underline{z} = \lambda \underline{z}$

set ~~[crossed out]~~ $\underline{\underline{A}} = \underline{\underline{I}} + r\underline{\underline{T}}$, where

$$\underline{\underline{T}} = \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & 1 & -2 & 1 \\ & & 1 & -2 \end{bmatrix} \longrightarrow (\underline{\underline{I}} + r\underline{\underline{T}})\underline{z} = \lambda \underline{z}$$

$$\longrightarrow r\underline{\underline{T}}\,\underline{z} = (\lambda - 1)\underline{z} \longrightarrow \cancel{\underline{\underline{T}}\,\underline{z} = \frac{\lambda-1}{r}\underline{z}}$$

$$\longrightarrow \boxed{\underline{\underline{T}}\,\underline{z} = \left(\frac{\lambda-1}{2}\right)\underline{z} \longrightarrow \underline{\underline{T}}\,\underline{z} = \nu \underline{z}, \quad \nu = \frac{\lambda-1}{r}}$$

eigenvalues of $\underline{\underline{T}}$:

$$\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \end{pmatrix} = \nu \begin{pmatrix} z_1 \\ z_2 \\ \vdots \end{pmatrix}$$

the k^{th} row of this system is:

$$z_{k-1} - 2z_k + z_{k+1} = \nu z_k \quad \text{for} \quad k = 1 \text{ to } N-1$$

assign $z_0 = z_N = 0$

$$\rightarrow \quad z_{k-1} - (2+\nu) z_k + z_{k+1} = 0$$

Assign $\quad 2 + \nu = 2\cos\Theta$

$$\rightarrow \quad z_{k-1} - 2\cos\Theta \, z_k + z_{k+1} = 0, \quad z_0 = z_N = 0 \rightarrow$$

Difference Equation with constant coeffs

Set $z_k = \xi^k$, ξ = constant, possibly complex
and ξ^k is the k^{th} power of ξ, not subscript

$$\rightarrow \quad \xi^{k-1} - 2\cos\Theta \, \xi^k + \xi^{k+1} = 0 \rightarrow$$

$$\xi^{k-1}\left(1 - 2\cos\Theta \, \xi + \xi^2\right) = 0$$

$$\xi = \frac{2\cos\Theta \pm \sqrt{4\cos^2\Theta - 4}}{2} = \frac{\cos\Theta \pm \sqrt{\cos^2\Theta - 1}}{1} = \cos\Theta \pm i\sin\Theta$$

$$\rightarrow \quad \xi = e^{\pm i\Theta}$$

$\rightarrow$ general solution: $z_k = c_1 e^{i\Theta k} + c_2 e^{-i\Theta k}$

set $z_0 = 0 \quad \rightarrow \quad c_1 = -c_2 \quad \rightarrow \quad z_k = c_1\left(e^{ik\Theta} - e^{-ik\Theta}\right)$

$$\rightarrow \quad z_k = 2ic_1 \sin(k\Theta)$$

set $z_N = 0 \quad \rightarrow \quad 2ic_1 \sin(N\Theta) = 0 \quad \rightarrow \quad N\Theta = p\pi$

$p = 1, 2, \ldots, N-1$

$$\rightarrow \quad \Theta = \frac{p\pi}{N}$$

work backwards:

$$2+\mu = 2\cos\Theta \;\longrightarrow\; \mu = -2(1-\cos\Theta)$$

$$\longrightarrow\; \mu = -2\left(1-\cos\frac{p\pi}{N}\right)$$

and now $\lambda = 1 + r\mu = 1 - 2r\left(1-\cos\frac{p\pi}{N}\right)$, $p = 1,\dots,N-1$

$$\longrightarrow\; \boxed{\lambda = 1-2r\left(1-\cos\frac{p\pi}{N}\right),\; p=1,\dots,N-1}$$

what is restriction on r so that $|\lambda| \le 1$???

$0 < 1-\cos\left(\frac{p\pi}{N}\right) < 2$, notice strict inequalities because $p = 1,\dots,N-1$

$$\longrightarrow\; 1-4r \ge -1 \;\longrightarrow\; \boxed{r \le \frac{1}{2}}$$

Finite Difference Methods for Parabolic Methods

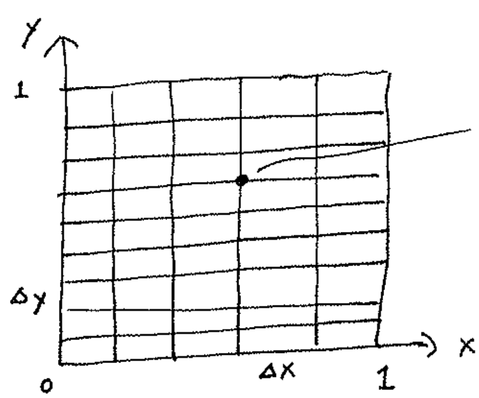
$u_t = \nu(u_{xx}+u_{yy})$, $0<x<1$, $0<y<1$

ICs: $u(x,y,0) = f(x,y)$, $t>0$

BCs: $u(x,y,t) = g(x,y,t)$ for (x,y) on the boundary

Approximate solution, grid (x_j, y_k), $j=0,\dots,N$, $k=0,\dots,M$

$x_j = j\Delta x$, $\Delta x = \frac{1}{N}$, $y_k = k\Delta y$, $\Delta y = \frac{1}{M}$



$v_{j,k}(t) \simeq u(x_j, y_k, t)$

Discrete equations: (method of lines)

$$v_{j,k}'(t) = \nu \left[\frac{1}{\Delta x^2} \delta_x^2 + \frac{1}{\Delta y^2} \delta_y^2 \right] v_{j,k}(t) \quad , \quad \begin{matrix} j = 1, \dots, N-1 \\ k = 1, \dots, M-1 \end{matrix}$$

$$v_{j,k}(t) = g(x_j, y_k, t) \quad \text{for } x_j, y_k \text{ on boundary}$$

SYSTEM OF ODES

Initial conditions, $v_{j,k}(0) = f(x_j, y_k)$

Now integrate in time. The simplest method of numerical integration is forward Euler:

$$\frac{v_{j,k}(t+\Delta t) - v_{j,k}(t)}{\Delta t} = \nu \left[\frac{1}{\Delta x^2} \delta_x^2 + \frac{1}{\Delta y^2} \delta_y^2 \right] v_{j,k}(t)$$

Set $t_n = n \Delta t$, $v_{j,k}(t_n) = v_{j,k}^n$

$$\boxed{\begin{gathered} v_{j,k}^{n+1} = v_{j,k}^n + \left[r_x \delta_x^2 v_{j,k}^n + r_y \delta_y^2 v_{j,k}^n \right] \\ r_x = \frac{\nu \Delta t}{\Delta x^2} \quad , \quad r_y = \frac{\nu \Delta t}{\Delta y^2} \end{gathered}}$$

explicit method advantage is that the implementation is straight forward

order of accuracy: $O(\Delta t, \Delta x^2, \Delta y^2)$

For stability, conduct a mode analysis:

$$v_{j,k}^n = a^n e^{i\alpha x_j} e^{i\beta y_k}$$

where α, β are wave numbers and a - growth factor

substitute into numerical scheme:

$$a = 1 + 2 r_x (\cos \alpha \Delta x - 1) + 2 r_y (\cos \beta \Delta y - 1)$$

$$a = 1 - 2 \left[r_x (1 - \cos \alpha \Delta x) + r_y (1 - \cos \beta \Delta y) \right]$$

we want to restrict r_x and r_y so that
$|a| \leq 1$ for $|\alpha \Delta x| \leq \pi$ and $|\beta \Delta y| \leq \pi$

when $\alpha \Delta x = \alpha \Delta y = 0$, $a = 1$,

we need $1 - 4(r_x + r_y) \geq -1 \rightarrow \frac{1}{2} \geq r_x + r_y$

$$\rightarrow \frac{\Delta t \nu}{\Delta x^2} + \frac{\Delta t \nu}{\Delta y^2} \leq \frac{1}{2} \rightarrow \Delta t \nu \left(\frac{\Delta x^2 + \Delta y^2}{\Delta x^2 \Delta y^2} \right) \leq \frac{1}{2}$$

$$\rightarrow \boxed{\Delta t \nu \leq \frac{1}{2} \frac{\Delta x^2 \Delta y^2}{\Delta x^2 + \Delta y^2}}$$ stability constraint

suppose $\Delta x = \Delta y = h$, then $\Delta t \nu \leq \frac{h^2}{4}$

this is a severe restriction on Δt

suggests the consideration of an implicit method
eg Backward Euler, Trapezoidal Rule, etc ....

9/29/03

Crank - Nicholson Method

$u_t = \nu(u_{xx} + u_{yy}) \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad t \geq 0$

$u(x,y,0) = f(x,y)$, $u(x,y,t) = g(x,y,t)$ on boundary

$$V_{j,k}'(t) = \nu \left[\frac{1}{\Delta x^2} \delta_x^2 + \frac{1}{\Delta y^2} \delta_y^2 \right] V_{j,k}(t) \rightarrow O(\Delta t, \Delta x^2, \Delta y^2)$$

For stability $\Delta t \leq \frac{1}{2\nu} \frac{\Delta x^2 \Delta y^2}{(\Delta x^2 + \Delta y^2)}$

For Forward Euler

IF we integrate using trapezoidal rule, obtain
Crank - Nicholson method

$$\frac{V_{j,k}^{n+1} - V_{j,k}^{n}}{\Delta t} = \nu \left[\frac{1}{\Delta x^2} \delta_x^2 + \frac{1}{\Delta y^2} \delta_y^2 \right] \left(\frac{V_{j,k}^{n+1} + V_{j,k}^{n}}{2} \right)$$

$$\frac{v_{j,k}^{n+1} - v_{j,k}^{n}}{\Delta t} = \nu \left[\frac{1}{\Delta x^2} \delta_x^2 + \frac{1}{\Delta y^2} \delta_y^2 \right] \left(\frac{v_{j,k}^{n+1} + v_{j,k}^{n}}{2} \right)$$

$$\rightarrow \quad v_{j,k}^{n+1} \left(1 - \frac{r_x}{2} \delta_x^2 - \frac{r_y}{2} \delta_y^2 \right) = \left(1 + \frac{r_x}{2} \delta_x^2 + \frac{r_y}{2} \delta_y^2 \right) v_{j,k}^{n}$$

where $r_x = \frac{\nu \Delta t}{\Delta x^2}$, $r_y = \frac{\nu \Delta t}{\Delta x^2}$

this is a linear system, $\boxed{\underline{\underline{A}}\, \underline{v}^{n+1} = \underline{c}^{n}}$

where $\underline{v}^{n} = \left[v_{1,1}^{n} \cdots v_{N-1,1}^{n} \,\vdots\, v_{1,2}^{n} \cdots v_{N-1,2}^{n} \,\vdots\, \cdots \,\vdots\, v_{1,M-1}^{n} \cdots v_{N-1,M-1}^{n} \right]^{T}$

$\underline{v}^{n}$ is an $(M-1)(N-1)$ vector

the matrix $\underline{\underline{A}}$ is a block tri-diagonal matrix

where each "sub" matrix is $(N-1) \times (M-1)$

$$\underline{\underline{A}} = \begin{bmatrix} \underline{\underline{A}}_{11} & \underline{\underline{A}}_{12} & & \\ \underline{\underline{A}}_{21} & \underline{\underline{A}}_{22} & \underline{\underline{A}}_{23} & \\ & \ddots & \ddots & \ddots \\ & & \underline{\underline{A}}_{N-2,M-2} & \underline{\underline{A}}_{N-1,M-1} \end{bmatrix}$$

where $$\underline{\underline{A}}_{kk} = \begin{bmatrix} 1+r_x+r_y & -\frac{r_x}{2} & & \\ -\frac{r_x}{2} & 1+r_x+r_y & -\frac{r_x}{2} & \\ & -\frac{r_x}{2} & 1+r_x+r_y & -\frac{r_x}{2} \\ & & \ddots & \ddots & \ddots \\ & & & -\frac{r_x}{2} & 1+r_x+r_y \end{bmatrix}$$

matrix on the diagonal

the matrices on sub-diagonal and super-diagonal

are $\underline{\underline{A}}_{k-1,k} = \underline{\underline{A}}_{k,k-1} = \frac{-r_y}{2} \underline{\underline{I}}$

$$\underline{c}^{n} = \begin{bmatrix} \underline{c}_1^{n} \\ \underline{c}_2^{n} \\ \vdots \\ \underline{c}_{N-1}^{n} \end{bmatrix} , \quad \underline{c}_k^{n} = \left[\left(1 + \frac{r_x}{2} \delta_x^2 + \frac{r_y}{2} \delta_y^2 \right) v_{j,k}^{n} \right]$$

the system is $(M-1)(M-1) \times (N-1)(N-1)$

boundary terms

when $k=1$:

$$\frac{r_y}{2}\begin{bmatrix} g(x_1, 0, (n+1)\Delta t) \\ \vdots \\ g(x_{N-1}, 0, (n+1)\Delta t) \end{bmatrix} + \frac{r_x}{2}\begin{bmatrix} g(0, y_1, (n+1)\Delta t) \\ \vdots \\ g(1, y_1, (n+1)\Delta t) \end{bmatrix}$$

when $k=2$:

$$\frac{r_y}{2}\begin{bmatrix} g(x_1, 0, (n+1)\Delta t) \\ \vdots \\ g(x_{N-1}, 0, (n+1)\Delta t \end{bmatrix} + \frac{r_x}{2}\begin{bmatrix} g(0, y_2, (n+1)\Delta t) \\ \vdots \\ g(0, y_2, (n+1)\Delta t) \end{bmatrix}$$

and so on for $k = 3, 4, \ldots$

Algorithm

1) specification $\nu \Delta t$, N, M, r_x, r_y
2) set $v^0_{j,k} = f(x_j, y_k)$
3) build $\underline{\underline{A}}$ (factor it)
4) for $n = 0, 1, \ldots, n_{final}$
   5) build $\underline{c}^n$ using $v^n_{j,k}$ + BCs
   6) solve $\underline{\underline{A}}\,\underline{v}^{n+1} = \underline{c}^n$
   7) construct $v^{n+1}_{j,k}$ from $\underline{v}^{n+1}$ + BCs
   8) output solution

Order of Accuracy

$$\tau^n_{j,k} = \nu\left[\frac{1}{\Delta x^2}\delta_x^2 + \frac{1}{\Delta y^2}\delta_y^2\right]\left(\frac{u^n_{j,k} + u^{n+1}_{j,k}}{2}\right) - \left(\frac{u^{n+1}_{j,k} - u_{j,k}}{\Delta t}\right)$$

the order is $O(\Delta t^2, \Delta x^2, \Delta y^2)$

Stability

set $v_{j,k}^{n} = a^{n} e^{i(\alpha x_j + \beta y_k)}$ wave numbers

$$\left[1 + r_x(1 - \cos\alpha\Delta x) + r_y(1 - \cos\beta\Delta y)\right] a$$
$$= \left[1 - r_x(1 - \cos\alpha\Delta x) - r_y(1 - \cos\beta\Delta y)\right]$$

$$a = \frac{1 - r_x(1 - z_1) - r_y(1 - z_2)}{1 + r_x(1 - z_1) + r_y(1 - z_2)} = \frac{N}{D}, \qquad \begin{array}{l} z_1 = \cos\alpha\Delta x \\ z_2 = \sin\beta\Delta y \end{array}$$

$$|a| < 1 \quad \Rightarrow \quad |N| \leq |D| \quad \Rightarrow \quad |N^2| \leq |D^2|$$

$$\cancel{1} + \cancel{r_x^2(1 - z_1)^2} + \cancel{r_y^2(1 - z_2)^2} - 2r_x(1 - z_1) - 2r_y(1 - z_2) + \cancel{2r_x r_y(1 - z_1)(1 - z_2)} \leq$$
$$\cancel{1} + \cancel{r_x^2(1 - z_1)^2} + \cancel{r_y^2(1 - z_2)^2} + 2r_x(1 - z_1) + 2r_y(1 - z_2) + \cancel{2r_x r_y(1 - z_1)(1 - z_2)}$$

$$\rightarrow \quad -r_x(1 - z_1) - r_y(1 - z_2) \leq r_x(1 - z_1) + r_y(1 - z_2)$$

$$\rightarrow \quad 0 \leq r_x(1 - z_1) + r_y(1 - z_2)$$

$$\left.\begin{array}{r} 1 - z_1 \geq 0 \\ 1 - z_2 \geq 0 \\ r_x \geq 0 \\ r_y \geq 0 \end{array}\right\} \text{unconditional stability}$$

$\Downarrow$

accuracy can now dictate the time step

Computational Cost

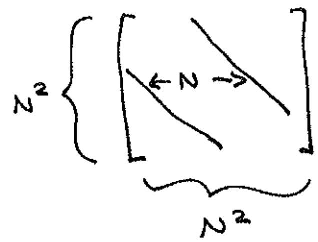
the main cost is in solving the linear system, but it is sparse and banded

IF we want a solution at an $O(1)$ time, $N \sim M \rightarrow O(N)$ time steps are required

total cost is $N \cdot$ (cost per step)

cost per step is based on the matrix solve of $\underline{\underline{A}}\, \underline{v}^{n+1} = \underline{c}^n$

assume, $N \sim M$:



direct solve, such as block tri-diagonal would require $O(\text{dim of system} \times \text{bandwidth}^2) = O(N^4)$ per step

$\Rightarrow$ cost $= O(N^5)$ total

<u>Iterative solvers:</u> conjugate gradients, SOR

$$O(\underbrace{\text{ops per sweep}}_{N^2} \times \underbrace{\#\text{ of sweeps}}_{N}) = O(N^3) \Rightarrow O(N^4) \text{ total}$$

some Fancy methods may achieve, $O(N^2 \log N)$
$\Rightarrow O(N^3 \log N)$ total

<u>Alternating Direction Implicit (ADI)</u>

commonly used for parabolic equations in 2, 3 dimensions

consider the heat equation

$$u_t = \nu(u_{xx} + u_{yy}), \quad 0 \le x \le 1, \quad 0 \le y \le 1, \quad t \ge 0$$

$$u(x,y,0) = f(x,y)$$

$$u(x,y,t) = g(x,y,t) \text{ on boundary}$$

the ADI method splits the implicit part into 2-half steps

The ADI scheme of Peaseman & Rachford

$$\frac{v_{j,k}^{n+1/2} - v_{j,k}^{n}}{\Delta t/2} = \nu\left[\frac{1}{\Delta x^2}\delta_x^2 v_{j,k}^{n+1/2} + \frac{1}{\Delta y^2}\delta_y^2 v_{j,k}^{n}\right]$$

$$\frac{v_{j,k}^{n+1} - v_{j,k}^{n+1/2}}{\Delta t/2} = \nu\left[\frac{1}{\Delta x^2}\delta_x^2 v_{j,k}^{n+1/2} + \frac{1}{\Delta y^2}\delta_y^n v_{j,k}^{n+1}\right]$$

boundary condition, $v_{j,k}^{n} = g(x_j, y_k, n\Delta t)$, (x_j, y_k) on boundary

initial condition $v_{j,k}^{0} = f(x_j, y_k)$

The advantage is that we only perform an implicit calculation in one dimension at a time, and each half step requires a tri-diagonal solve.

<u>First half-step:</u>

$$v_{j,k}^{n+1/2} - v_{j,k}^{n} = r_x\delta_x^2 v_{j,k}^{n+1/2} + r_y\delta_y^2 v_{j,k}^{n} \quad , \; r_x = \frac{\nu\Delta t}{2\Delta x^2} \; , \; r_y = \frac{\nu\Delta t}{2\Delta y^2}$$

$$\rightarrow \left(1 - r_x\delta_x^2\right) v_{j,k}^{n+1/2} = \left(1 + r_y\delta_y^2\right) v_{j,k}^{n}$$

$$\underline{v}_k^n = \begin{bmatrix} v_{1,k}^{n+1/2} \\ \vdots \\ v_{N-1,k}^{n+1/2} \end{bmatrix} , \quad \underline{\underline{A}}_n = \begin{bmatrix} 1+2r_x & r_x & & \\ -r_x & 1+2r_x & -r_x & \\ & \ddots & \ddots & \\ & & -r_x & 1+2r_x \end{bmatrix}$$

$$\underline{c}_k^n = \begin{bmatrix} \left(1 + r_y\delta_y^2\right)v_{1,k}^{n} + r_x\, g(x_0, y_k, t_{n+1/2}) \\ \left(1 + r_y\delta_y^2\right)v_{2,k}^{n} \\ \vdots \\ \left(1 + r_y\delta_y^2\right)v_{N-1,k}^{n} + r_x\, g(x_N, y_k, t_{n+1/2}) \end{bmatrix}$$

the first half-step becomes:

$$\underline{\underline{A}}_k \underline{v}_k^n = \underline{c}_k^n \quad , \; k = 1, \ldots, M-1$$

$\llcorner$ tridiagonal $\rightarrow O(NM)$ ops

Similarly, the second half step

$\underline{\underline{A}}_k \underline{v}_j^n = \underline{c}_j^n \quad , \quad j = 1, \ldots,$

$\uparrow O(NM)$, so each step requires $O(NM)$ operations

Consistency and Order of Accuracy & Stability

1^{st} step: $(1 - r_x \delta_x^2) v_{j,k}^{n+1/2} = (1 + r_y \delta_y^2) v_{j,k}^n$

2^{nd} step: $(1 - r_y \delta_y^2) v_{j,k}^{n+1} = (1 + r_x \delta_x^2) v_{j,k}^{n+1/2}$

Note: $(1 - r_x \delta_x^2)(1 - r_y \delta_y^2) v_{j,k}^{n+1} = (1 - r_x \delta_x^2)(1 + r_x \delta_x^2) v_{j,k}^{n+1/2}$

$$= (1 + r_x \delta_x^2)(1 + r_y \delta_x^2) v_{j,k}^n$$

$$\rightarrow \Delta t\, \tau_{j,k}^n = (1 - r_x \delta_x^2)(1 - r_y \delta_y^2) u_{j,k}^{n+1} - (1 + r_x \delta_x^2)(1 + r_y \delta_y^2) u_{jk}^n$$

note, $r_x \delta_x^2 u_{jk}^n = r_x (\Delta x^2 u_{xx} + O(\Delta x^4))$

$$= \frac{\nu \Delta t}{2} u_{xx} + O(\Delta x^2)$$

$$r_y \delta_y^2 u_{jk}^n = \frac{\nu \Delta t}{2} u_{yy} + O(\Delta y^2)$$

$$\rightarrow \Delta t\, \tau_{j,k}^n = \left[u - \frac{\nu \Delta t}{2}(u_{xx} + u_{yy}) + \frac{\nu^2 \Delta t^2}{4} u_{xx} u_{yy} + O(\Delta x^2 \Delta t, \Delta y^2 \Delta t) \right]^{n+1} - \left[u + \frac{\nu \Delta t}{2}(u_{xx} + u_{yy}) + \frac{\nu^2 \Delta t^2}{4} u_{xxyy} + O(\Delta x^2 \Delta t, \Delta y^2 \Delta t) \right]^n$$

$$\rightarrow \Delta t\, \tau_{jk}^n = \Delta t\, u_t + \frac{\Delta t^2}{2} u_{tt} + O(\Delta t^3) - \nu \Delta t (u_{xx} + u_{yy}) - \frac{\nu \Delta t^2}{2}(u_{xx} + u_{yy})_t + O(\Delta t^3) + \frac{\nu^2 \Delta t^3}{4} u_{xxyyt} + O(\Delta x^2 \Delta t, \Delta y^2 \Delta t)$$

$$\boxed{\tau_{jk}^n = \cancel{\Delta t}\left(O(\Delta t^2, \Delta x^2, \Delta y^2) \right)}$$

Methods for Heat Equation in Multi-space dimensions

for example, $u_t = \nu(u_{xx} + u_{yy})$, $0<x<1$, $0<y<1$

$u(x,y,0) = f(x,y)$

u is given on boundary

the Method of Lines leads to a whole family of methods

need a grid: $x_j = j\Delta x$, $\Delta x = \frac{1}{N}$

$y_k = k\Delta y$, $\Delta y = \frac{1}{M}$

$$v_{j,k}(t) = u(x_j, y_k, t)$$

Let $v_{j,k}$ solve the ODEs $v_{j,k}'(t) = \nu\left(\frac{1}{\Delta x^2}\delta_x^2 + \frac{1}{\Delta y^2}\delta_y^2\right)v_{j,k}$

for $1 \le j \le N-1$, $1 \le k \le M-1$, with initial condition $v_{j,k}(0) = f(x_j,$

and boundary conditions given at $j=0,N$, $k=0,M$

<u>Time integration</u>

a) Forward Euler: • $\tau_{j,k}^n = O(\Delta t, \Delta x^2, \Delta y^2)$ accuracy

- stability restriction: $\nu\frac{\Delta t}{\Delta x^2} + \frac{\nu\Delta t}{\Delta y^2} \le 1 \leftarrow$ restrictive because explicit method
- easy to code, but time step would necessarily be too small.

b) Trapezoidal Rule $\Rightarrow$ Crank-Nicholson Method

- $\tau_{j,k}^n = O(\Delta t^2, \Delta x^2, \Delta y^2)$ accuracy
- unconditionally stable ("can take any Δt you want, and scheme will be stable, but not necessarily accurate, so take Δt on order of $\Delta x, \Delta y$")
- implicit method so a little harder to code, (must solve linear system)

c) ADI, Peaseman - Rachford

- 2-level alternating implicit scheme
- $\tau_{j,k}^{n} = O(\Delta t^2, \Delta x^2, \Delta y^2)$
- unconditionaly stable
- moderately hard to code
- choose $\Delta t \sim \Delta x \sim \Delta y$ to balance

~~First order operator~~ δ_x

$$\frac{v_{j,k}^{*} - v_{j,k}^{n}}{\Delta t/2} = \nu \left[\frac{1}{\Delta x^2} \delta_x^2 v_{j,k}^{*} + \frac{1}{\Delta y^2} \delta_y^2 v_{j,k}^{n} \right]$$

implicit   explicit

$$\frac{v_{j,k}^{n+1} - v_{j,k}^{*}}{\Delta t/2} = \nu \left[\frac{1}{\Delta x^2} \delta_x^2 v_{j,k}^{*} + \frac{1}{\Delta y^2} \delta_y^2 v_{j,k}^{n+1} \right]$$

explicit   implicit

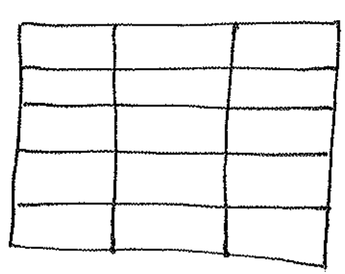
You can show that this is 2^{nd} order accurate in both time and space.

ADI Algorithm Structure

1) Initialization - grid, time-step, and other parameters...
   set initial grid function

2) Main time-stepping loop

   3) for $n = 0, \ldots, n_{Final}$



$\uparrow k$

$\rightarrow j$

$O(MN)$

4) for $k = 1, \ldots, M-1$

   5) build tri-diagonal matrix system, solve it, save solution on grid line — $O(N)$

- end

we have $v_{j,k}^{*}$, don't need $v_{j,k}^{n}$

$O(MN)$ $\left\{\begin{array}{l} \text{6) For } j = 1, \ldots, N-1 \\ O(M) \left\{\begin{array}{l} \text{7) build tridiagonal matrix, solve it,} \\ \quad \text{save solution on grid line} \end{array}\right. \\ \bullet \text{ end} \end{array}\right.$

you have $v^{n+1}_{j,k}$ and don't need $v^{*}_{j,k}$

7) output results

8) end time loop

$\Rightarrow$ each time step costs $O(MN)$

$\Rightarrow$ total cost is $O(MN)$ per time step

$\Rightarrow$ this is optimal

Polar Coordinates

the domain is now a disk, $0 < r < 1$, $0 \leq \theta \leq 2\pi$, $t > 0$

$$u_t = \frac{1}{r}(r u_r)_r + \frac{1}{r^2} u_{\theta\theta}$$

where the diffusivity, $\nu = 1$

initial condition: $u(r, \theta, 0) = f(r, \theta)$

boundary condition: $u(1, \theta, t) = g(\theta, t)$

there are additional boundary conditions that are implied by this problem

$$u(r, \theta, t) = u(r, \theta + 2\pi, t) \quad (2\pi \text{ periodic})$$

the singularity at $r = 0$ is "ok"

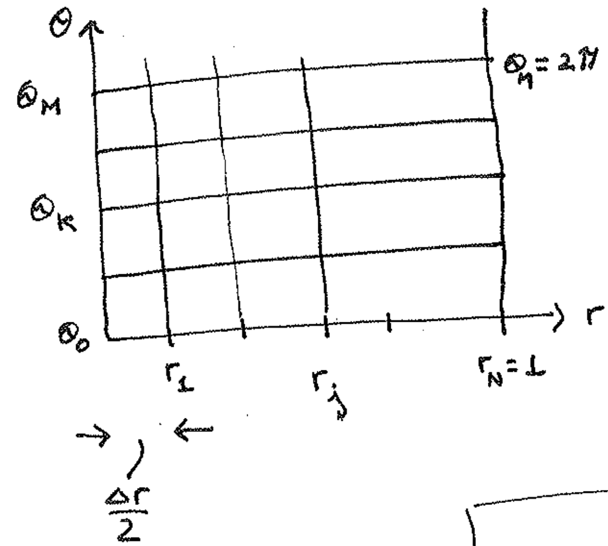
then enforce $\lim_{r \to 0} r u_r = 0$

<u>grid</u>: $r_j = \left(j - \frac{1}{2}\right)\Delta r$, want $r_N = 1$

$\rightarrow$ solve for Δr, $1 = (N - \tfrac{1}{2})\Delta r \rightarrow \boxed{\Delta r = \frac{1}{N - \tfrac{1}{2}}}$

$\theta_k = k\Delta\theta$, want $\theta_M = 2\pi \rightarrow \boxed{\Delta\theta = \frac{2\pi}{M}}$

<u>polar grid</u>:



to approximate $\frac{1}{r}(r u_r)_r$, use $\frac{1}{r}\left[\frac{(r u_r)_{j+1/2} - (r u_r)_{j-1/2}}{\Delta r}\right]$

let $v_{j,k}(t) \approx u(r_j, \theta_k, t)$

$$\rightarrow \quad v_{j,k}'(t) = \frac{1}{r_j \Delta r^2}\left[r_{j+1/2}\,\delta_{+r} - r_{j-1/2}\,\delta_{-r}\right] v_{j,k} + \frac{1}{r_j^2}\frac{1}{\Delta\theta^2}\,\delta_\theta^2 v_{j,k}$$

where $v_{j,k}(0) = f(r_j, \theta_k)$

and $v_{N,k}(t) = g(\theta_k, t)$

and implied BCs:
- periodicity, since $v_{j,0}(t) = v_{j,M}(t)$ then solve ODEs on grid for $k = 1, \ldots, M$
  likewise, $\boxed{v_{j,M+1}(t) = v_{j,1}(t)}$
- at $j = 1$, forward difference term is ok, but backward difference term will have $r_{j-1/2}$

$$\boxed{r_{j-1/2}\,\delta_{-r} v_{j,k} = 0 \quad \text{for } j = 1}$$

Time-integration

want to handle Θ-differences implicitly because of $\frac{1}{r_j^2}$ coefficient

consider the following scheme:

$$\frac{v_{j,k}^{n+1} - v_{j,k}^{n}}{\Delta t} = \frac{1}{\Delta r^2} D\, v_{j,k}^{n} + \frac{1}{r_j^2}\cdot\frac{1}{\Delta\Theta^2}\,\delta_\Theta^2\, v_{j,k}^{n+1}$$

where $D = \frac{1}{r_j}\left[r_{j+1/2}\,\delta_{+r} - r_{j-1/2}\,\delta_{-r} \right]$

combining terms, obtain

$$\left[1 - \frac{\Delta t}{r_j^2\,\Delta\Theta^2}\,\delta_\Theta^2 \right] v_{j,k}^{n+1} = \left(1 + \frac{\Delta t}{\Delta r^2} D \right) v_{j,k}^{n}$$

consider a fixed j grid line

$$\begin{bmatrix} \times & \times & & & \times \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \ddots & \ddots & \ddots \\ \times & & & \times & \times \end{bmatrix} \begin{pmatrix} v_{j,1}^{n+1} \\ v_{j,2}^{n+1} \\ \vdots \\ v_{j,M}^{n+1} \end{pmatrix} =$$

due to periodicity

to solve this system, partition the matrix into actual tridiagonal matrix and border parts

$$\left[\begin{array}{cccccc:c} \times & \times & & & & & \times \\ \times & \times & \times & & & & \\ & \times & \times & \times & & & \\ & & \times & \times & \times & & \\ & & & \times & \times & & \times \\ \hdashline \times & & & & \times & & \times \end{array}\right]$$

$\boxed{\text{Bordering Algorithms}}$

denote this matrix as

$$\left[\begin{array}{c:c} \underline{\underline{A}} & \underline{b} \\ \hdashline \underline{c}^T & d \end{array}\right] \begin{pmatrix} \underline{x} \\ \hdashline y \end{pmatrix} = \begin{pmatrix} \underline{r} \\ \hdashline s \end{pmatrix}$$

$$\left[\begin{array}{c:c} \underline{\underline{A}} & \underline{b} \\ \hdashline \underline{c}^T & d \end{array}\right] \left(\begin{array}{c} \underline{x} \\ \hdashline y \end{array}\right) = \left(\begin{array}{c} \underline{r} \\ \hdashline s \end{array}\right)$$, assume that $\underline{\underline{A}}$ is nonsingular

solve $\underline{\underline{A}}\,\underline{p} = \underline{r}$, $\underline{\underline{A}}\,\underline{q} = \underline{b}$ $\Rightarrow$ $\underline{p}, \underline{q}$ are known

$$\left[\begin{array}{c:c} \underline{\underline{A}}^{-1} & \underline{0} \\ \hdashline \underline{0} & 1 \end{array}\right] \left[\begin{array}{c:c} \underline{\underline{A}} & \underline{b} \\ \hdashline \underline{c}^T & d \end{array}\right] \left(\begin{array}{c} \underline{x} \\ \hdashline y \end{array}\right) = \left[\begin{array}{c:c} \underline{\underline{A}}^{-1} & \underline{0} \\ \hdashline \underline{0} & 1 \end{array}\right] \left(\begin{array}{c} \underline{r} \\ \hdashline s \end{array}\right)$$

$$\Rightarrow \quad \left[\begin{array}{c:c} \underline{\underline{I}} & \underline{q} \\ \hdashline \underline{c}^T & d \end{array}\right] \left(\begin{array}{c} \underline{x} \\ \hdashline y \end{array}\right) = \left(\begin{array}{c} \underline{p} \\ \hdashline s \end{array}\right)$$

now multiply all of this times $\left[\begin{array}{c:c} \underline{\underline{I}} & \underline{0} \\ \hdashline -\underline{c}^T & 1 \end{array}\right]$ $\Rightarrow$ $\left[\begin{array}{c:c} \underline{\underline{I}} & \underline{0} \\ \hdashline -\underline{c}^T & 1 \end{array}\right] \left[\begin{array}{c:c} \underline{\underline{I}} & \underline{q} \\ \hdashline \underline{c}^T & d \end{array}\right]$

$$\Rightarrow \quad \left[\begin{array}{c:c} \underline{\underline{I}} & \underline{q} \\ \hdashline \underline{0} & \tilde{d} \end{array}\right] \left(\begin{array}{c} \underline{x} \\ \hdashline y \end{array}\right) = \left(\begin{array}{c} \underline{p} \\ \hdashline \tilde{s} \end{array}\right)$$ where $\tilde{d} = d - \underline{c}^T \underline{q}$

$\tilde{s} = s - \underline{c}^T \underline{p}$

$\Rightarrow$ results : $y = \dfrac{\tilde{s}}{\tilde{d}}$, $\underline{x} = \underline{p} - y\,\underline{q}$

A nonlinear Problem for the Heat Eqn

$u_t = u_{xx} + b(u)$, $0<x<1$, $t>0$

where $b(u)$ is some nonlinear source

initial condition , $u(x,0) = f(x)$

boundary conditions : $u(0,t) = u(1,t) = 0$

Finite Difference approximation.

grid: $x_j = j\Delta x$, $\Delta x = \frac{1}{N}$, $v_j(t) \cong u(x_j, t)$ - method of lines

Let $v_j(t)$ solve

$$v_j'(t) = \frac{1}{\Delta x^2}\delta_x^2 v_j + b(v_j^n) \quad , 1 \le j \le N-1 \quad \text{(interior grid lines)}$$

$$v_j(0) = f(x_j) \quad , \quad v_0(t) = v_N(t) = 0$$

We want to consider a 2nd order implicit integration

$$\boxed{\frac{v_j^{n+1} - v_j^n}{\Delta t} = \frac{1}{\Delta x^2}\delta_x^2\left(\frac{v_j^n + v_j^{n+1}}{2}\right) + b\left(\frac{v_j^n + v_j^{n+1}}{2}\right) \quad , 1 \le j \le N, n \ge 0 \qquad \text{"Fully discrete difference scheme"} \qquad v_j^0 = f(x_j) \quad , \quad v_0^n = v_N^n = 0}$$

The question is how do we solve for v_j^{n+1}

$$F_j \equiv v_j^{n+1} - v_j^n - r\delta_x^2\left(\frac{v_j^n + v_j^{n+1}}{2}\right) - \Delta t\, b\left(\frac{v_j^n + v_j^{n+1}}{2}\right) = 0 \quad , \quad r = \frac{\Delta t}{\Delta x^2}$$

this is a system of nonlinear equations :

$$\underline{F}(\underline{v}) = 0 \quad , \quad \underline{v} = \left[v_1^{n+1} \quad v_2^{n+1} \quad \cdots \quad v_{N-1}^{n+1}\right]^T$$

(N-1 nonlinear algebraic equations)

use Newton's method to solve

Initial guess: $\underline{v}^{(0)} = \underline{v}$ at the n^{th} time level

For $p = 0, 1, \ldots$

solve $\underline{F}_{\underline{v}}(\underline{v}^{(p)})\, \underline{dv}^{(p)} = \underline{F}(\underline{v}^{(p)})$ where $\underline{F}_{\underline{v}}$ is the Jacobian

update $v^{(p+1)} = v^{(p)} - \underline{dv}^{(p)}$

stop when $\| \underline{dv}^{(p)} \| \leq tol \cdot \| \underline{v}^{(p)} \|$ where $tol \sim 10^{-8}$

$\boxed{\text{if } \underline{F}_{\underline{v}} \text{ is nonsingular at root and initial guess sufficiently close, then convergence is quadratic}}$

Form of $\underline{F}_{\underline{v}}$:

$$F_1 = v_1^{n+1} - v_1^n - \frac{r}{2}\Big(v_2^{n+1} - 2v_1^{n+1} + \underbrace{\cancel{v_0^{n+1}}}_{\text{known}} \Big)^{0,\,BC} - \frac{r}{2}\,\underbrace{\delta_x^2 v_1^n}_{\text{known}} - \Delta t\, b\, b\left(\frac{v_1^{n+1} + v_1^n}{2} \right)$$

$$\frac{\partial F_1}{\partial v_1} = 1 + r - \Delta t\, b' \left(\frac{v_1^{n+1} + v_1^n}{2} \right) \cdot \frac{1}{2}$$

$$\frac{\partial F_1}{\partial v_2} = \frac{-r}{2} \quad , \quad \frac{\partial F_1}{\partial v_j} = 0\,,\ j \geq 3$$

similarly

$$\frac{\partial F_2}{\partial v_1} = -\frac{r}{2} \quad , \quad \frac{\partial F_2}{\partial v_2} = 1 + r - \Delta t\, b'\left(\frac{v_1^{n+1} + v_1^n}{2} \right) \cdot \frac{1}{2} \quad , \quad \frac{\partial F_2}{\partial v_3} = \frac{-r}{2} \quad , \quad \frac{\partial F_2}{\partial v_j} = 0\,,\ j \geq 4$$

therefore $\underline{F}_{\underline{v}}$ is tridiagonal

$$\frac{\partial F}{\partial v} = \begin{bmatrix} \beta_1 & \gamma_1 & & & \\ \alpha_2 & \beta_2 & \gamma_2 & & \\ & \alpha_3 & \beta_3 & \gamma_3 & \\ & & \ddots & \ddots & \ddots \\ & & & & \end{bmatrix}$$

where $\alpha_j = \gamma_j = \frac{-r}{2}$

$$\beta_j = 1 + r - \frac{\Delta t}{2}\, b'\left(\frac{v_j^{n+1} + v_j^n}{2} \right)$$

Convergence, consistency, stability

Error: $w_j^n = v_j^n - u_j^n$ where $u_j^n = u(x_j, t_n)$

substitute $v_j^n = w_j^n + u_j^n$ into difference scheme

$$u_j^{n+1} + w_j^{n+1} - (u_j^n + w_j^n) = r\,\delta_x^2\left(\frac{u_j^n + u_j^{n+1}}{2}\right) + r\,\delta_x^2\left(\frac{w_j^n + w_j^{n+1}}{2}\right)$$

$$+ \Delta t\, \underbrace{b\left(\frac{u_j^n + u_j^{n+1}}{2} + \frac{w_j^n + w_j^{n+1}}{2}\right)}_{\text{expand about } b\left(\frac{u_j^n + u_j^{n+1}}{2}\right)}$$

$$\to \quad b\left(\frac{u_j^n + u_j^{n+1}}{2} + \frac{w_j^n + w_j^{n+1}}{2}\right) = b\left(\frac{u_j^n + u_j^{n+1}}{2}\right) + b'(\Theta_j^n)\,\frac{w_j^n + w_j^{n+1}}{2}$$

$\to$ Taylor series, one term, with remainder

where Θ_j^n is between $\frac{u_j^n + u_j^{n+1}}{2}$ and $\frac{u_j^n + u_j^{n+1}}{2} + \frac{w_j^n + w_j^{n+1}}{2}$

Truncation error:

$$\Delta t\, \tau_j^n = u_j^{n+1} - u_j^n - r\,\delta_x^2\left(\frac{u_j^n + u_j^{n+1}}{2}\right) - \Delta t\, b\left(\frac{u_j^n + u_j^{n+1}}{2}\right), \quad r = \frac{\Delta t}{\Delta x^2}$$

$$= \Delta t\, u_t + \frac{\Delta t^2}{2} u_{tt} + O(\Delta t^3) - \frac{\Delta t}{2}\Big[u_{xx} + \frac{\Delta x^2}{12} u_{xxxx} + O(\Delta x^4)$$

$$+ u_{xx} + \Delta t\, u_{xxt} + O(\Delta t^2) + \frac{\Delta x^2}{12} u_{xxxx} + O(\Delta x^2 \Delta t)\Big]$$

$$- \Delta t\, b\left[u + \frac{\Delta t}{2} u_t + O(\Delta t^2)\right]$$

Note, $b\left(u + \frac{\Delta t}{2} u_t + \dots\right) = b(u) + b'(u)\frac{\Delta t}{2} u_t + \dots$

$$\Rightarrow \quad \tau_j^n = \cancel{u_t} + \cancel{\frac{\Delta t}{2} u_{tt}} + O(\Delta t^2) \cancel{- u_{xx}} - \cancel{\frac{\Delta t}{2} u_{xxt}} + O(\Delta x^2)$$

$$\cancel{-b(u)} - \cancel{\frac{\Delta t}{2} b'(u) u_t} + O(\Delta t^2)$$

from PDE

time derivative of PDE

$$\Rightarrow \quad \boxed{\tau_j^n = O(\Delta t^2) + O(\Delta x^2)}$$

So back to the error equation

$$w_j^{n+1} = w_j^n + r\delta_x^2\left(\frac{w_j^n + w_j^{n+1}}{2}\right) + \Delta t\, b'(\Theta_j^n)\left(\frac{w_j^n + w_j^{n+1}}{2}\right) - \Delta t\, \tau_j^n$$

to analyze this error equation, need to freeze $b'(\Theta_j^n)$ to be some constant and consider the stability of the resulting linear constant coeff equation for all possible values for $\sigma = b'(\Theta_j^n)$

$\Rightarrow$ necessary condition for stability (in order to obtain sufficient condition, need to do nonlinear analysis)

Midterm

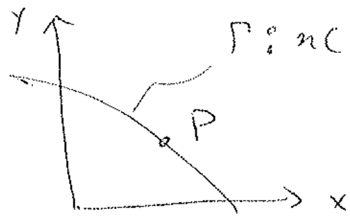
closed book, closed notes, 1 crib sheet

1) classification of PDEs

$$A u_{xx} + 2B u_{xy} + C u_{yy} = D$$

nonlinear PDE, but linear in highest-derivatives so we call it quasi-linear

issue of classification boils down to being able to construct equation locally



can you construct solution u locally about point P, given u and u_n on Γ at P

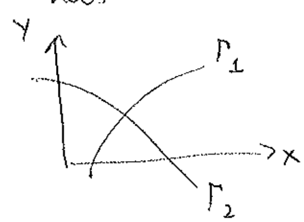
Cases

1) $B^2 - AC > 0$ - hyperbolic, may or may not be able to construct (form Taylor series) solution u depending on whether Γ is a characteristic

differential form of characteristic equation

$$A\,dy^2 - 2B\,dx\,dy + C\,dx^2 = 0$$

has real solutions



issues include domain of dependence region of influence

2) $B^2 - AC < 0$ - elliptic - no real characteristics, solutions of equations are analytic

3) $B^2 - AC = 0$, parabolic case - may or may not be able to build solution locally

- **Canonical Forms**

$u_{xx} + u_{yy} = \delta$, elliptic

$u_{xy} = \delta$, hyperbolic

$u_{xx} = \delta$. , parabolic

where δ only depends on lower derivatives

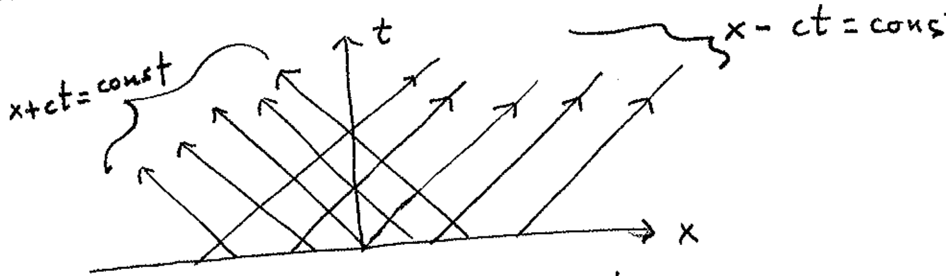
- Boundary conditions for well-posed problems

2) Finite Difference Methods
- heat eqn w/ Dirichlet BCs
- heat eqn w/ Neumann BCs
- ghost lines, staggered grids
- generating discrete approximation
    - Taylor series approach
    - interpolation approach
    - finite-volumes, discrete conservation
- Theoretical issues, truncation error
    - consistency, order of accuracy,
    - convergence, global error
    - stability - do errors in difference equation grow?
      well-posedness of the difference equation
    - Lax theorems
      if consistent
- stability analysis
    - Fourier stability analysis, Discrete Fourier Transform
    - Fourier mode analysis
    - Matrix analysis, if boundaries included
- Parabolic Eqns
    - Crank-Nicholson method
    - ADI
    - Polar coordinates
    - Non linear Eqn

Hyperbolic PDEs

A hyperbolic PDE is one that possesses a full set of real characteristics.

Example, wave equation, $u_{tt} = c^2 u_{xx}$. This PDE has two families of real characteristics, $x - ct =$ constant, $x + ct =$ constant.



Can also consider the wave equation as a system of two first-order equations.

$$\underline{u}_t + \underline{\underline{A}}\, \underline{u}_x = 0 \qquad\qquad u^{(1)}_t + c\, u^{(2)}_x = 0 \quad (1)$$

$$\text{where} \quad \underline{u} = \begin{bmatrix} u^{(1)} \\ u^{(2)} \end{bmatrix}, \quad \underline{\underline{A}} = \begin{bmatrix} 0 & c \\ c & c \end{bmatrix} \quad \Rightarrow \quad u^{(2)}_t + c\, u^{(1)}_x = 0 \quad (2)$$

Take $\frac{\partial}{\partial t}$ of eqn (1) and subtract $c \frac{\partial}{\partial x}$ of eqn (2) $\rightarrow u^{(1)}_{tt} - c^2 u^{(1)}_{xx} = 0$

Therefore the first component of $\underline{u}$ satisfies the wave equation. Similarly, you can show that the second component also satisfies the wave equation.

Consider the linear system of first order PDEs

$$\underline{u}_t + \underline{\underline{A}}\, u_x = 0, \qquad \underline{\underline{A}} \in \mathbb{R}^{m \times m}, \quad \underline{u}(x,t) = [u^{(1)}\ u^{(2)} \ldots u^{(m)}]^T$$

This system is called hyperbolic if A is diagonalizable with real eigenvalues

Assume that $\underline{\underline{A}}$ is diagonalizable with real eigenvalues.

Then a nonsingular matrix $\underline{\underline{R}}$ exists such that

$\underline{\underline{R}}^{-1}\underline{\underline{A}}\,\underline{\underline{R}} = \underline{\underline{\Lambda}}$ where $\underline{\underline{\Lambda}} = \text{diag}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix}$, a diagonal matrix

with real eigenvalues of $\underline{\underline{A}}$ on diagonals.

Set $\underline{w}(x,t) = \underline{\underline{R}}^{-1}\underline{u}(x,t)$ = "characteristic variables", "Riemann variables"

Note : rows of $\underline{\underline{R}}^{-1}$ are left eigenvectors

and columns of $\underline{\underline{R}}$ are the right eigenvectors

$$\rightarrow R^{-1}\left(\underline{u}_t + \underline{\underline{A}}\,\underset{RR^{-1}}{\uparrow}\,\underline{u}_x = 0\right) \rightarrow \underline{\underline{R}}^{-1}\underline{u}_t + \underline{\underline{R}}^{-1}\underline{\underline{A}}\,\underline{\underline{R}}\,\underline{\underline{R}}^{-1}\underline{u}_x = 0$$

$$\rightarrow (\underline{\underline{R}}^{-1}\underline{u})_t + \underline{\underline{\Lambda}}(\underline{\underline{R}}^{-1}\underline{u})_x = 0 \rightarrow \underline{w}_t + \underline{\underline{\Lambda}}\,\underline{w}_x = 0$$

$$\rightarrow \frac{\partial}{\partial t}\begin{bmatrix} w^{(1)} \\ w^{(2)} \\ \vdots \\ w^{(m)} \end{bmatrix} + \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix} w^{(1)} \\ w^{(2)} \\ \vdots \\ w^{(m)} \end{bmatrix}$$

the equations have decoupled, the p^{th} equation is

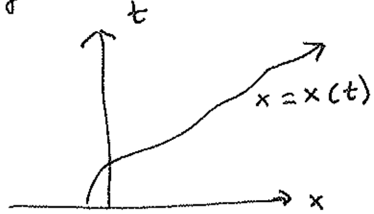
$$w_t^{(p)} + \lambda_p w_x^{(p)} = 0, \quad p = 1,\dots,m$$

$\rightarrow$ each characteristic variable satisfies a scalar linear advection equation

Linear Advection Equation

$u_t + c u_x = 0$, $c \in \mathbb{R}$, $|x| < \infty$, $t > 0$

Solve using the method of characteristics. Consider the rate of change of u with respect to t along a path $x = x(t)$



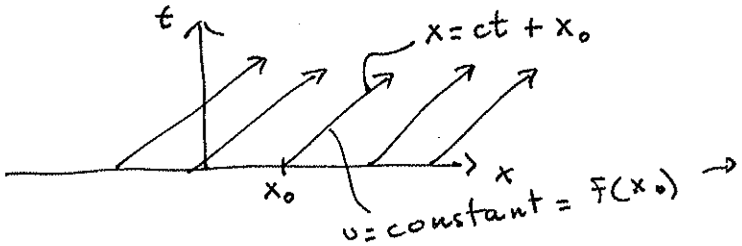
$$\frac{d}{dt}\, u(x(t),t) = \frac{\partial u}{\partial x}\frac{dx}{dt} + \frac{\partial u}{\partial t}$$

select $\frac{dx}{dt} = c$ then $\frac{d}{dt}u(x(t),t) = 0$

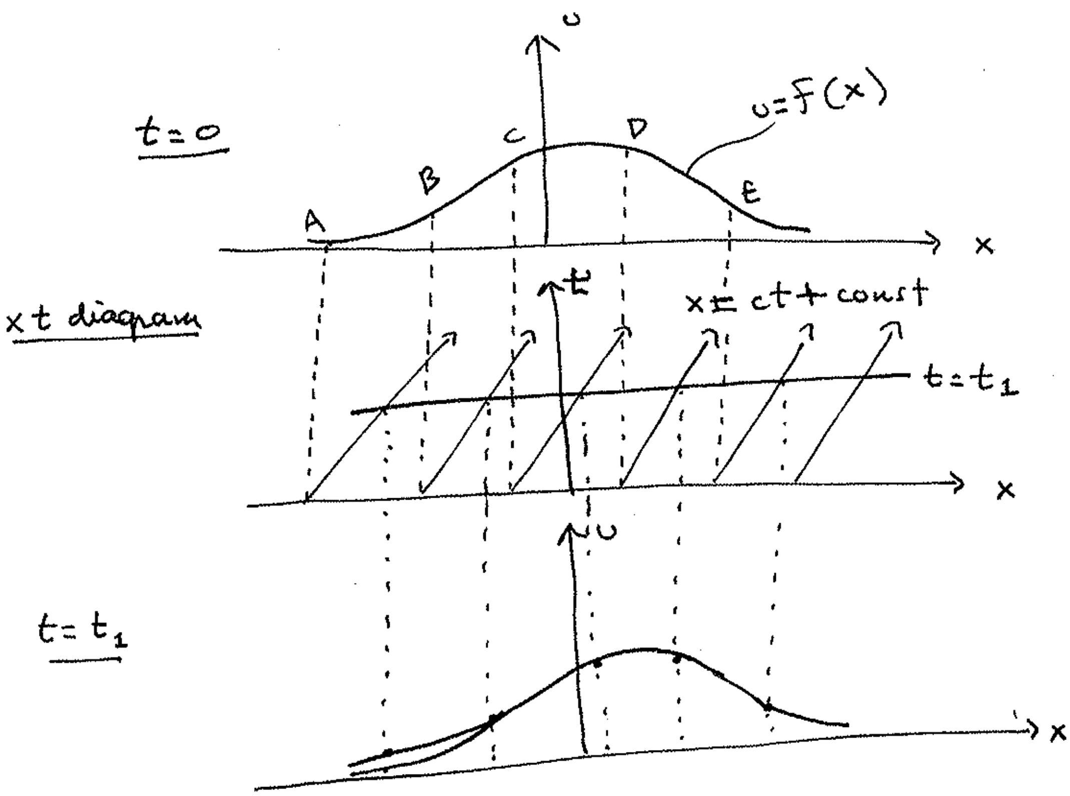
$\rightarrow$ characteristic equations are $\frac{du}{dt} = 0$ along $\frac{dx}{dt} = c$

solve the characteristic equations

$\rightarrow$ $u =$ constant along $x = ct +$ constant



$\rightarrow$ $\boxed{u(x,t) = f(x-ct)}$



the solution translates to the right at speed $c > 0$ and unchanged shape

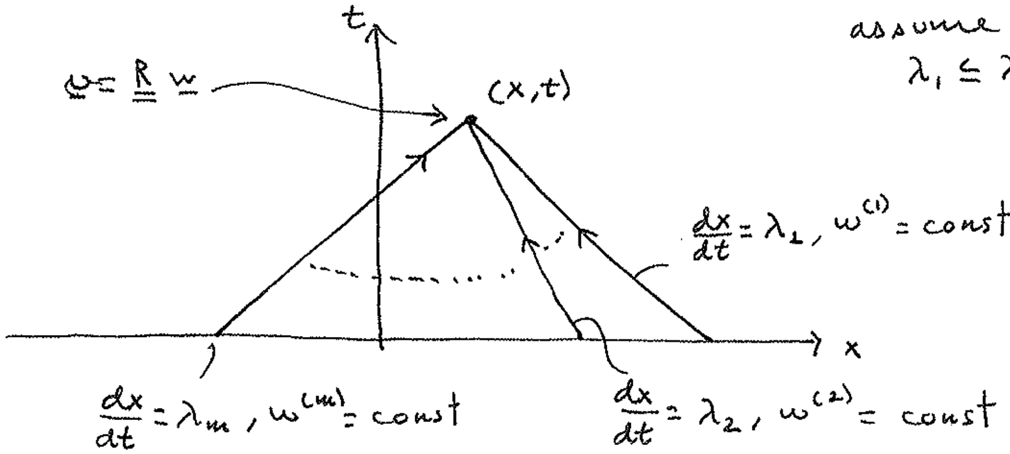
Return to the linear system

$$\frac{\partial}{\partial t} w^{(p)} + \lambda_p \frac{\partial}{\partial x} w^{(p)} = 0 \quad , \quad p = 1, 2, \dots, m$$

solution

$$w^{(p)} = \text{constant along } x = \lambda_p t + \text{constant}$$

assume
$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$



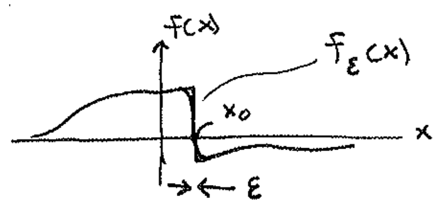
if $\underline{u}(x,0)$ is given, then $\underline{w}(x,0) = \underline{\underline{R}}^{-1} \underline{u}(x,0)$,
the characterist variables initial values are known

Behavior of Discontinuities for Linear Equations

$$u_t + c u_x = 0 \quad , \quad |x| < \infty , \quad t > 0 \quad , \quad u(x,0) = f(x)$$

solution: $u(x,t) = f(x-ct)$

Suppose f has the form:

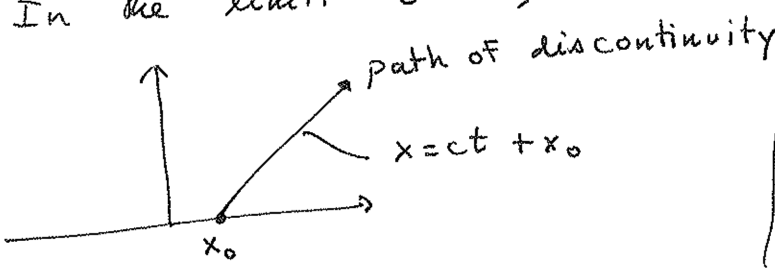


In the limit $\varepsilon \to 0$, $f(x)$ becomes discontinuous at x_0

so $f_\varepsilon(x)$ is smooth if $\varepsilon > 0$

$\to \quad u(x,t) = f_\varepsilon(x-ct) \quad , \quad \varepsilon > 0$

In the limit $\epsilon \to 0$, $u(x,t) = f(x-ct)$



For a linear problem, the discontinuities are a result of a discontinuity in the initial condition

Numerical Methods

scalar advection equation, $u_t + cu_x = 0$, $|x| < \infty$, $t > 0$, $f(x) = u(x,0)$

grid: $x_j = j\Delta x$, Δx = given, > 0

$t_n = n\Delta t$, Δt = given, > 0

let $v_j^n \simeq u(x_j, t_n)$

standard methods:

- "centered" methods:

Lax-Friedrichs method:

$$v_j^{n+1} = \frac{1}{2}\left(v_{j-1}^n + v_{j+1}^n\right) - \frac{c\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right)$$

Lax-Wendroff method:

$$v_j^{n+1} = v_j^n - \frac{c\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right) + \frac{1}{2}\left(\frac{c\Delta t}{\Delta x}\right)^2\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n\right)$$

- "one-sided" methods (Upwind methods)

$$v_j^{n+1} = v_j^n - \frac{c\Delta t}{\Delta x}\left(v_j^n - v_{j-1}^n\right)$$

$$v_j^{n+1} = v_j^n - \frac{c\Delta t}{\Delta x}\left(v_{j+1}^n - v_j^n\right)$$

- Remarks
  1) Most methods for hyperbolic equations are <u>explicit</u> because the stability constraints for these methods are $\Delta t \le \text{const}\cdot\Delta x$ which is probably what you would pick for accuracy
  2) First choice might be a Method of Lines construction

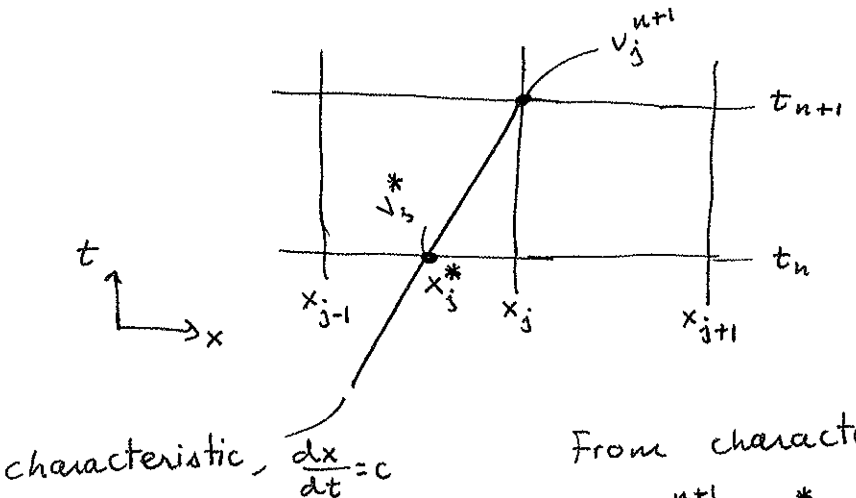
set $v_j(t) \simeq u(x_j, t)$ $\rightarrow$ $v_j'(t) + c\underbrace{\frac{v_{j+1} - v_{j-1}}{2\Delta x}}_{\substack{2^{nd}\text{ order centered} \\ \text{approximation for } u_x}} = 0$

then use Forward Euler to integrate:

$$\frac{v_j^{n+1} - v_j^n}{\Delta t} + \frac{c}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right) = 0$$

$$\rightarrow \quad v_j^{n+1} = v_j^n - \frac{c\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right) \rightarrow \text{UNCONDITIONALLY UNSTABLE}$$

3) Previous methods can be derived by consideration of characteristics



From characteristic construction,

$$v_j^{n+1} = v_j^* \qquad (\text{exact})$$

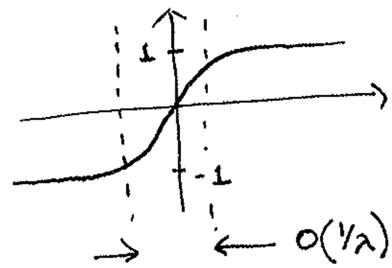
Approximate value of v_j^* using grid data at t_n via some form of <u>interpolation</u>:

- <u>linear interpolation</u> yields the <u>upwind method</u>
  consider data points (x_{j-1}, v_{j-1}^n) and (x_j, v_j^n)
  or perhaps data points (x_j, v_j^n) and (x_{j+1}, v_{j+1}^n) $(1^{st}$ order$)$
- Interpolate v_j^* using a <u>linear fit</u> to
  (x_{j-1}^n, v_{j-1}^n), (x_{j+1}^n, v_{j+1}^n) $\Rightarrow$ <u>Lax-Friedrichs</u> $(1^{st}$ order$)$
- Interpolate v_j^* using a quadratic fit to
  (x_{j-k}^n, v_{j-k}^n) for $k = -1, 0, 1$ $\Rightarrow$ <u>Lax-Wendroff</u>
  2^{nd} order accurate

Example: linear advection equaton

$$u_t + c u_x = 0 \ , \quad t > 0 \ ,$$

$$u(x,0) = \tanh \lambda x \quad , \ \lambda = \text{const.}$$



solution: $u(x,t) = \tanh \lambda (x - ct)$

the parameter λ determines the transition width, $O(1/\lambda)$

Compare with various numerical approximations for different values of Δx and λ.

1st order scheme: Lax-Friedrichs

$$v_j^{n+1} = \frac{1}{2}\left(v_{j-1}^n + v_{j+1}^n\right) - \frac{c\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right)$$

2nd order scheme: Lax-Wendroff

$$v_j^{n+1} = v_j^n - \frac{c\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right) + \frac{1}{2}\left(\frac{c\Delta t}{\Delta x}\right)^2\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n\right)$$

Compare and contrast these two schemes.

Integrate from $t=0$ to $t=5$

with $\Delta t = \sigma \Delta x$ where $\sigma \simeq 1$

(Note, $\sigma = 1$ is the stability limit, where both methods become exact.)

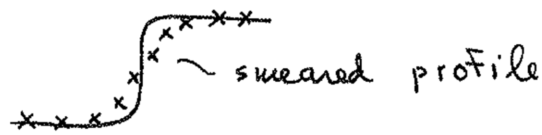
Observations:

1) when λ is small and the solution is well represented on the grid, then

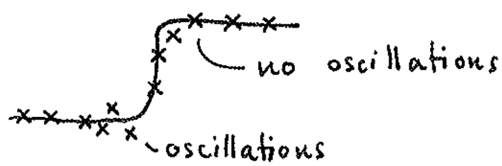
$$\left.\begin{array}{l} \text{error } (LF) \text{ is } O(\Delta t, \Delta x) \\ \text{error } (LW) \text{ is } O(\Delta t^2, \Delta x^2) \end{array}\right\} \text{ expected behavior}$$

2) when λ is large and the solution is not well represented on the grid, then we observe significant error near the transition

LF, the error is smooth



LW, the error shows oscillations



3) If $|\sigma| = 1 \Rightarrow$ numerical solution becomes exact for this simple equation (because the characteristic is exactly aligned with the grid)

Analyze The Behavior of the Error For Both Methods

Lax-Friedrichs: $v_j^{n+1} = \frac{1}{2}\left(v_{j-1}^n + v_{j+1}^n\right) - \frac{c\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right)$

solve for truncation error, τ_j^n:

$$u_j^{n+1} = \frac{1}{2}\left(u_{j-1}^n + u_{j+1}^n\right) - \frac{c\Delta t}{2\Delta x}\left(u_{j+1}^n - u_{j-1}^n\right) + \Delta t\, \tau_j^n$$

$$\rightarrow u + \Delta t\, u_t + \frac{\Delta t^2}{2} u_{tt} + \cdots = \frac{1}{2}\left(2u + \Delta x^2 u_{xx} + \cdots\right)$$

$$- \frac{c\Delta t}{2\Delta x}\left(2\Delta x\, u_x + O(\Delta x^3)\right) + \Delta t\, \tau_j^n$$

Divide through by Δt

$$u_t + \frac{\Delta t}{2} u_{tt} + O(\Delta t^2) = \frac{\Delta x^2}{2\Delta t} u_{xx} + O(\Delta x^4/\Delta t) - c u_x + O(\Delta x^2) + \tau$$

- If we assign $u(x,t)$ as a solution of $u_t + c u_x = 0$, then the truncation error, $\tau = \left(\Delta t, \frac{\Delta x^2}{\Delta t}\right)$ = first order accurate provided $\frac{\Delta x}{\Delta t}$ = constant, which it is.
- If we assign $u(x,t)$ as a solution of the <u>modified equation</u>

$$u_t + c u_x = -\frac{\Delta t}{2} u_{tt} + \frac{\Delta x^2}{2\Delta t} u_{xx}$$

then the truncation error is smaller. In other words, the numerical solution approximates the solution of the modified equation better than the original.

$$u_t = - c u_x + \cdots \quad \leftarrow \text{modified equations}$$

$$\rightarrow u_{tt} = - c u_{xt} + \cdots = -c(u_t)_x + \cdots = -c(-c u_x)_x + \cdots$$

$$\rightarrow u_{tt} = c^2 u_{xx} + \cdots$$

higher order in $\Delta t, \Delta x$ so <u>ignored</u>

$$\Rightarrow \quad u_t + c u_x = -\frac{\Delta t}{2}(c^2 u_{xx} + \cdots) + \frac{\Delta x^2}{2\Delta t} u_{xx}$$

<u>Modified equation (to leading order)</u>

$$u_t + c u_x = \frac{\Delta t}{2}\left(\frac{\Delta x^2}{\Delta t^2} - c^2\right) u_{xx}$$

$$\rightarrow \quad u_t + c u_x = \frac{c^2 \Delta t}{2}\left(\frac{\Delta x^2}{c^2 \Delta t^2} - 1\right) u_{xx}$$

$$\rightarrow \quad \boxed{u_t + c u_x = \frac{c^2 \Delta t}{2}\left(\frac{1}{r^2} - 1\right) u_{xx}, \quad r = \frac{c\Delta t}{\Delta x}}$$

The general form of the modified equation is thus $u_t + c u_x = \nu u_{xx}$, where $\nu = \frac{c^2 \Delta t}{2}\left(\frac{1}{r^2} - 1\right)$

The term νu_{xx} provides the leading behavior of the error in Lax-Friedrichs. It is a diffusive term, and $\boxed{\text{therefore the leading behavior of error is diffusive (or dissipative) for Lax-Friedrichs}}$

If we consider ν to be a diffusivity, then we'll need $\nu \geq 0$, otherwise end up with ill-posed backwards heat equation, so a stability constraint falls out of the analysis.

Fundamental solutions

let $u(x,t) = e^{\lambda t + ikx}$. Substitute into modified equation:

$$\lambda e^{\lambda t + ikx} + ikc\, e^{\lambda t + ikx} = -\nu k^2 e^{\lambda t + ikx}$$

$$\rightarrow \lambda = -ikc - \nu k^2 \quad \Rightarrow \quad \boxed{u(x,t) = e^{ik(x-ct) - \nu k^2 t}}$$

the $e^{ik(x-ct)}$ represents advection whereas as the $e^{-\nu k^2}$ represents a small amplitude decay

general solution : $u(x,t) = \int_{-\infty}^{\infty} c(k)\, e^{ik(x-ct) - \nu k^2 t}\, dk$

Behavior of the Error for Numerical Methods of the Linear Advection Equation

$u_t + u_x = 0$, $(c = 1)$, $u(x,0) = f(x) = \tanh \lambda x$

exact solution $u(x,t) = f(x - ct)$

Lax-Friedrichs (typical first order method)

$$v_j^{n+1} = \frac{1}{2}\left(v_{j-1}^n + v_{j+1}^n\right) - \frac{\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right)$$

One method of studying error is to derive the modified equation. The idea is to replace the leading order term of the truncation error as a forcing term.

Modified Equation:

$$u_j^{n+1} = \frac{1}{2}\left(u_{j-1}^n + u_{j+1}^n\right) - \frac{\Delta t}{2\Delta x}\left(u_{j+1}^n - u_{j-1}^n\right) + \Delta t\, \tau_j^n$$

Expand in Taylor series and retain the leading order terms in τ:

$$\underbrace{u_t + u_x = \nu u_{xx}}_{\text{advection - diffusion equation}}, \quad \nu = \frac{\Delta t}{2}\left(\frac{1}{\sigma^2} - 1\right), \quad \sigma = \frac{\Delta t}{\Delta x}$$

the leading behavior of the error is diffusive

Fundamental solution of modified eqn, $u = e^{\lambda t + ikx}$ (growth factor: λ; wave number: k)

$$\rightarrow \quad u_k = e^{ik(x-t) - \nu k^2 t}$$

by superposition,

$$u(x,t) = \int_{-\infty}^{\infty} c(k)\, e^{ik(x-t) - \nu k^2 t}\, dk$$

where $c(k)$ determined by initial solution

At $t=0$, $u(x,0) = \int_{-\infty}^{\infty} c(k) e^{ikx}\, dk = F(x)$

The question is, when does the term $e^{-\nu k^2 t}$ contribute?

The error is related to the size of $\nu k^2 t$.

If our initial state $F(x)$ is well represented on the grid, then the coefficient $|c(k)|$ would be small when $|k\Delta x|$ is large $\Rightarrow$ error is small.

$$e^{ik(x-t)-\nu k^2 t} = e^{k(i(x-t)-\nu k t)} \quad , \nu = \frac{\Delta t}{2}\left(\frac{1}{\sigma^2}-1\right), \quad \sigma = \frac{\Delta t}{\Delta x}$$

All of this is true for t sufficiently small.

If $F(x)$ is not well represented on the grid then $|c(k)|$ is large when $|k\Delta x|$ is large $\Rightarrow$ significant error.

Lax-Wendroff Methods

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{2\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right) + \frac{1}{2}\left(\frac{\Delta t}{\Delta x}\right)^2\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n\right)$$

with initial condition $v_j^0 = F(x_j)$

Modified equation: $u_t + u_x = \nu u_{xxx}$, $\nu = -\frac{\Delta t^2}{6}\left(\frac{1}{\sigma^2}-1\right)$, $\sigma = \frac{\Delta t}{\Delta x}$

the dominant behavior of the error is no longer diffusive

$u_t + u_x = \nu u_{xxx} \rightarrow$ linear KDV equation, arises in shall water waves

can derive solution by separation of variables

— Dispersive effect —

$u_t + u_x = \nu u_{xxx}$

substitute $u = e^{\lambda t + ikx}$

$\to \lambda = -ik - i\nu k^3 \to u_k = e^{-ikt - i\nu k^3 t + ikx}$

$\to u_k = e^{ik(x - (t + \nu k^2 t))} \to u_k = e^{ik(x - (1 + \nu k^2) t)}$

phase error

$\to$ general solution: $\boxed{u(x,t) = \int_{-\infty}^{\infty} c(k)\, e^{ik(x - (1+\nu k^2) t)}\, dk}$, $\nu = -\frac{\Delta t^2}{6}\left(\frac{1}{\sigma^2} - 1\right)$

IF $F(x)$ is well represented on grid then $|c(k)|$ is small when $|k\Delta x|$ is large $\Rightarrow$ ~~phase errors~~ no sign error. IF $F(x)$ is not well represented on the grid then the magnitude of c, $|c|$ is large when $|k\Delta x|$ is large $\Rightarrow$ sign phase error, dispersive.

The next term in the truncation error is a dissipation term, so there is some dissipation, but at high order.

Stability and the CFL condition

(Courant-Friedrichs-Lewy)

Look at one sided methods for $u_t + c u_x = 0$.

(1) $v_j^{n+1} = v_j^n - \sigma(v_j^n - v_{j-1}^n)$, $\sigma = \frac{c\Delta t}{\Delta x}$, Backwards DiFF

(2) $v_j^{n+1} = v_j^n - \sigma(v_{j+1}^n - v_j^n)$, Forward DiFF

Stability analysis, $v_j^n = a^n e^{ikx_j}$

(1) $\Rightarrow a = 1 - \sigma(1 - e^{-ik\Delta x})$

(2) $\Rightarrow a = 1 - \sigma(e^{ik\Delta x} - 1)$

(1) : $\quad a = 1 - \sigma\left(1 - e^{-ik\Delta x}\right)$

(2) : $\quad a = 1 - \sigma\left(e^{ik\Delta x} - 1\right)$

(1) $\Rightarrow \quad |a|^2 = \left(1 - \sigma(1 - \cos k\Delta x)\right)^2 + \sigma^2 \sin^2 k\Delta x \le 1$

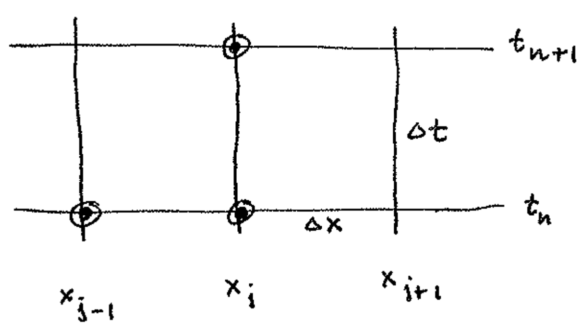
$$\Rightarrow \quad \sigma(1-\sigma) \ge 0 \quad \Rightarrow \quad \underline{0 \le \sigma \le 1 \ \text{ for stability}}$$

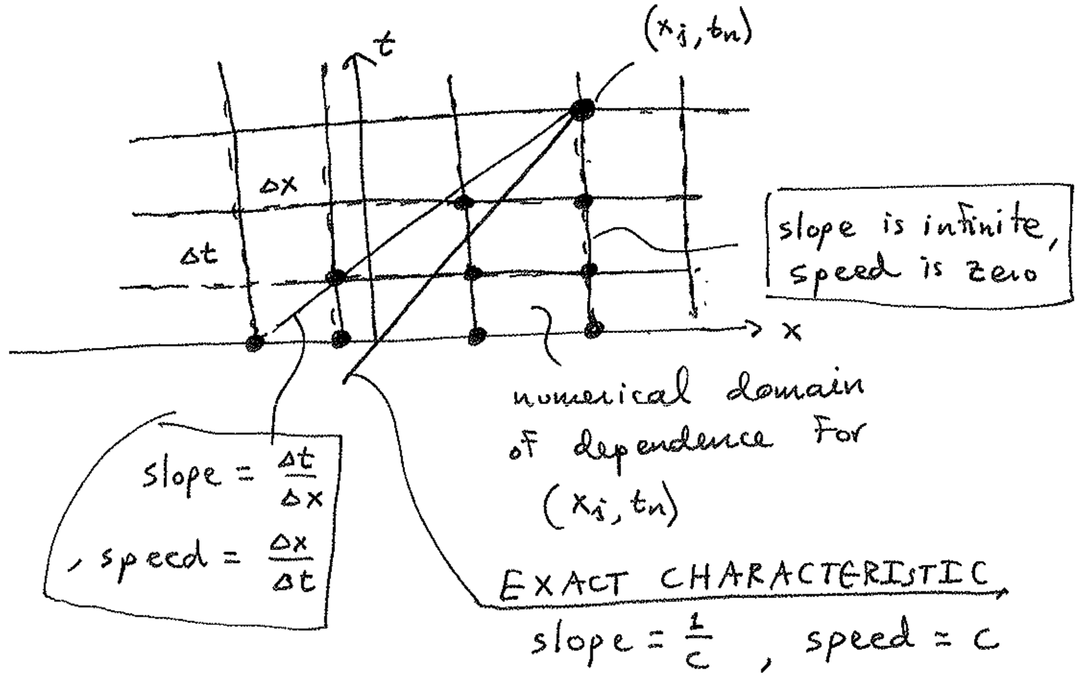
(2) $\Rightarrow \quad |a|^2 = \left(1 + \sigma(1 - \cos k\Delta x)\right)^2 + \sigma^2 \sin^2 k\Delta x \le 1$

$$\Rightarrow \quad \underline{-1 \le \sigma \le 0 \ \text{ for stability}}$$

We can interpret these results in terms of the domain of dependence.

scheme (1) Backward Difference, $v_j^{n+1} = v_j^n - \sigma\left(v_j^n - v_{j-1}^n\right)$



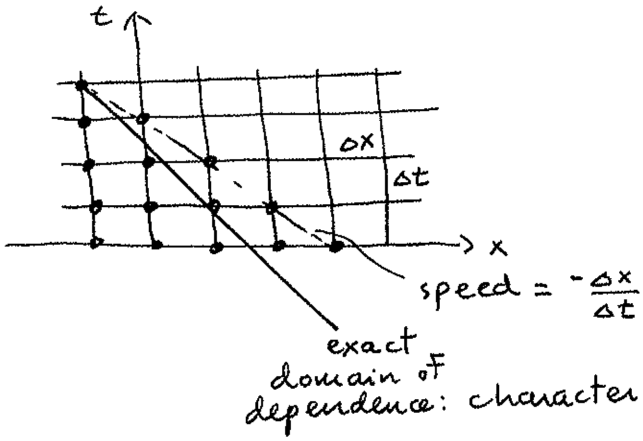


Stability result for (1), backwards diff

$\Rightarrow \quad 0 \le \sigma \le 1 \quad \Rightarrow \quad 0 \le \frac{c\Delta t}{\Delta x} \le 1 \quad \Rightarrow \quad 0 \le c \le \frac{\Delta x}{\Delta t}$

$\Rightarrow$ the exact domain of dependence must be contained within the numerical domain of dependence $\Rightarrow$ CFL condition (necessary condition for stability)

Similarly, for (2), Forwards Difference



$\Rightarrow \quad -1 \le \sigma \le 0 \quad \Rightarrow \quad -\frac{\Delta x}{\Delta t} \le c \le 0$

Non-Constant Coeffs

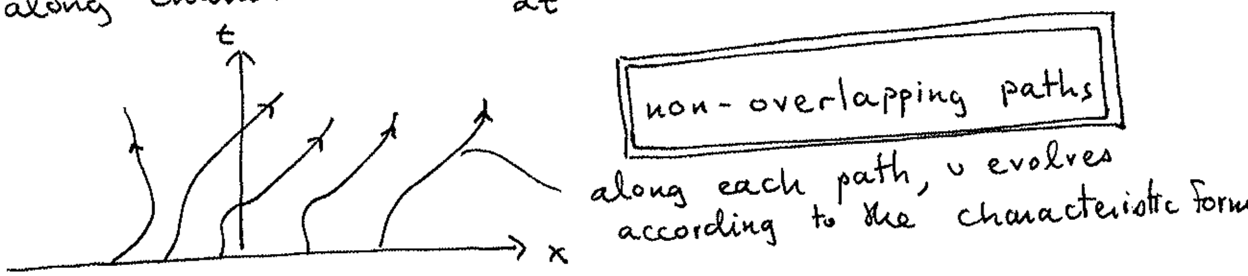
General <u>linear</u> scalar eqn:

$$u_t + c(x,t)\, u_x = a(x,t)\, u + b(x,t)$$

Characteristic form:

$$\frac{du}{dt} = a\big(x(t),t\big)\, u + b\big(x(t),t\big), \quad u(x,0) = u(x_0,0)$$

along characteristics $\frac{dx}{dt} = c(x,t)$, $x(0) = x_0$



c)

Numerical solution

$$u_t + c(x,t)\,u_x = a(x,t)\,u + b(x,t)$$

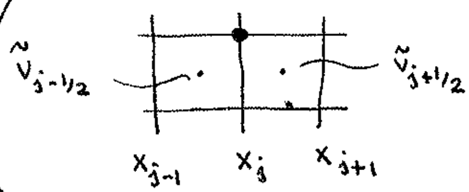
◆ lower order method, such as **upwind**:

$$v_j^{n+1} = v_j^n - \frac{\Delta t\, c(x_j,t_n)}{\Delta x}\left(v_j^n - v_{j-1}^n\right) + \Delta t\left(a(x_j,t_n)\,v_j^n + b(x_j,t^n)\right)$$

stability implies that $0 \le c(x_j,t_n) \le \frac{\Delta x}{\Delta t}$

◆ 2^{nd} order scheme: 2-step Lax-Wendroff

For simple advection equation, $u_t + c u_x = 0$



$$\tilde{v}_{j+1/2} = \frac{1}{2}\left(v_j^n + v_{j+1}^n\right) - \frac{c\Delta t}{2\Delta x}\left(v_{j+1}^n - v_j^n\right)$$

$$v_j^{n+1} = v_j^n - \frac{\Delta t c}{\Delta x}\left(\tilde{v}_{j+1/2} - \tilde{v}_{j-1/2}\right)$$

2^{nd} order accurate

For the general, linear, non-constant coeff eqn:

$$\tilde{v}_{j+1/2} = \frac{1}{2}\left(v_j^n + v_{j+1}^n\right) + \frac{c(x_{j+1/2},t_n)\Delta t}{2\Delta x}\left(v_{j+1}^n - v_j^n\right) + \frac{\Delta t}{2}\Big(a(x_{j+1/2},t_n)\frac{1}{2}\left(v_j^n + v_{j+1}^n\right) + b(x_{i+1/2},t_n)\Big)$$

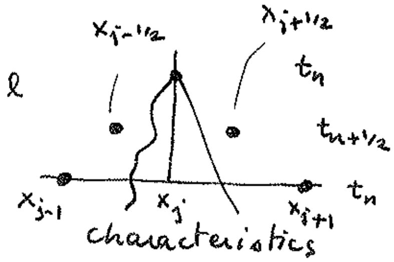
$$v_j^{n+1} = v_j^n - \frac{\Delta t\; c(x_j,t_{n+1/2})}{\Delta x}\left(\tilde{v}_{j+1/2} - \tilde{v}_{j-1/2}\right) + \Delta t\Big(a(x_j,t_{n+1/2})\frac{1}{2}\left(\tilde{v}_{j-1/2} + \tilde{v}_{j+1/2}\right) + b(x_j,t_{n+1/2})\Big)$$

stability implies

must worry about fastest wave speed

$$\max_{j,n} |c(x_j,t_n)|\,\frac{\Delta t}{\Delta x} \le 1$$

as can be seen by stencil



Systems of Linear Equations

$$\underline{u}_t + \underline{\underline{A}}\,\underline{u}_x = 0 \quad , \quad \underline{\underline{A}} \in \mathbb{R}^{m\times m}$$

The system is hyperbolic if $\underline{\underline{A}}$ is diagonalizable and with real eigenvalues.

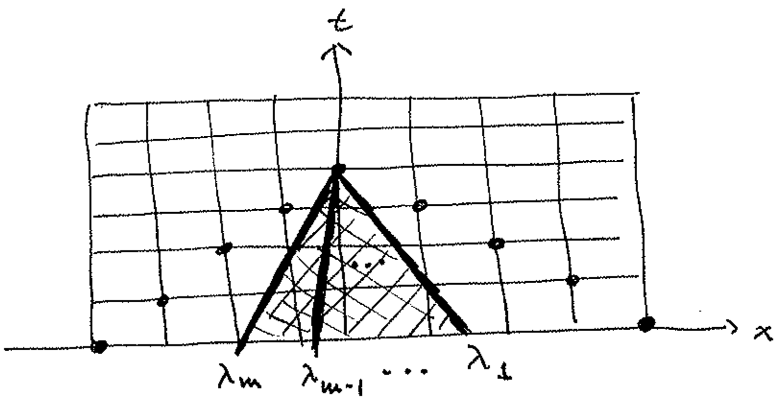
Lax-Friedrichs:

$$\underline{v}_j^{n+1} = \frac{1}{2}\left(\underline{v}_{j-1}^n + \underline{v}_{j+1}^n\right) - \underline{\underline{A}}\frac{\Delta t}{2\Delta x}\left(\underline{v}_{j+1}^n - \underline{v}_{j-1}^n\right)$$

Lax-Wendroff:

$$\underline{v}_j^{n+1} = \underline{v}_j^n - \underline{\underline{A}}\frac{\Delta t}{2\Delta x}\left(\underline{v}_{j+1}^n - \underline{v}_{j-1}^n\right) + \underline{\underline{A}}^2\left(\frac{\Delta t}{\Delta x}\right)^2\cdot\frac{1}{2}\left(\underline{v}_{j+1}^n - 2\underline{v}_j^n + \underline{v}_{j-1}^n\right)$$

CFL Stability Analysis



Suppose $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_m$ are the eigenvalues of $\underline{\underline{A}}$

$$\max_{1\le p\le m} |\lambda_p| \frac{\Delta t}{\Delta x} \le 1$$

$$\Rightarrow \Delta t \le \frac{\Delta x}{\max_{1\le p\le m}|\lambda_p|}$$

Faster the propagation, smaller the time step

d)

Upwind Methods (have less diffusion than Lax-Friedrichs)

$\underline{u}_t + \underline{\underline{A}}\,\underline{u}_x = 0$,

suppose the eigenvalues of $\underline{\underline{A}}$ are

$$\lambda_1 \le \cdots \le \lambda_q \le 0 \le \lambda_{q+1} \le \cdots \le \lambda_m$$

the sign of the eigenvalues determines direction of discrete derivatives

$$\underline{\underline{R}}^{-1}\underline{u}_t + \underline{\underline{R}}^{-1}\underline{\underline{A}}\,\underline{\underline{R}}\,\underline{\underline{R}}^{-1}\underline{u}_x = 0\,, \qquad \underline{\underline{R}}^{-1}\underline{u}_t + \underline{\underline{\Lambda}}\,\underline{\underline{R}}^{-1}\underline{u}_x = 0$$

$\rightarrow \underline{w} = \underline{\underline{R}}^{-1}\underline{u}$, $\quad \rightarrow \underline{w}_t + \underline{\underline{\Lambda}}\,\underline{w}_x = 0$

set

$$\underline{\underline{\Lambda}} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_q & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{bmatrix}}_{\underline{\underline{\Lambda}}_-} + \underbrace{\begin{bmatrix} 0 & & & & \\ & \ddots\, 0 & & & \\ & & \lambda_{q+1} & & \\ & & & \ddots & \\ & & & & \lambda_m \end{bmatrix}}_{\underline{\underline{\Lambda}}_+}$$

$$\rightarrow \quad \underline{w}_t + (\underline{\underline{\Lambda}}_- + \underline{\underline{\Lambda}}_+) w_x = 0$$

multiply through by $\underline{\underline{R}}$ to recover $\underline{u}$

$$\underline{u}_t + \underbrace{R\Lambda_- R^{-1}}_{A_-} u_x + \underbrace{R\Lambda_+ R^{-1}}_{A_+} u_x = 0 \qquad \text{with } A = A_- + A_+$$

$\rightarrow \quad u_t + A_- u_x + A_+ u_x = 0$ decomposed eqn into parts with left running characteristics and right running characteristics

the upwind discretization

$$\underline{v}_j^{n+1} = \underline{v}_j^n - \frac{\Delta t}{\Delta x}\underline{\underline{A}}_-\left(\underline{v}_{j+1}^n - \underline{v}_j^n\right) - \frac{\Delta t}{\Delta x}\underline{\underline{A}}_+\left(\underline{v}_j^n - \underline{v}_{j-1}^n\right)$$

$$= \underline{v}_j^n - \frac{\Delta t}{2\Delta x}\underline{\underline{A}}\left(\underline{v}_{j+1}^n - \underline{v}_{j-1}^n\right) + \frac{\Delta t}{2\Delta x}|\underline{\underline{A}}|\left(\underline{v}_{j+1}^n - 2\underline{v}_j^n + \underline{v}_{j-1}^n\right)$$

where $|\underline{\underline{A}}| = \underline{\underline{A}}_+ - \underline{\underline{A}}_-$

Boundary Conditions for Linear Hyperbolic Systems

$$\underline{u}_t + \underline{\underline{A}}\,\underline{u}_x = 0 \,, \quad 0<x<1, \quad t>0 \quad , \underline{u}(x,0) = \underline{f}(x)$$

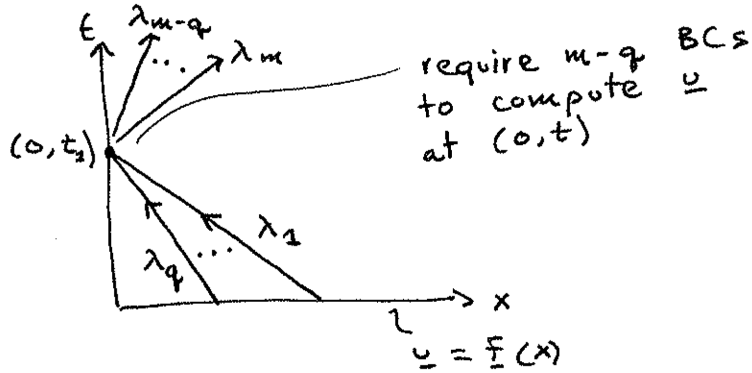
What boundary conditions are allowed?

We must consider the characteristics:

$\underline{\underline{A}}$ has eigenvalues $\lambda_1, \dots, \lambda_m$

where $\lambda_1 \le \lambda_2 \le \dots \le \lambda_q < 0 < \lambda_{q+1} \le \dots \le \lambda_m$

Near the boundary $x = 0$:



the characteristics $\lambda_1, \dots, \lambda_q \to$ outflow characteristics

the characteristics $\lambda_{q+1}, \dots, \lambda_m$ are inflow characteristics

Rule: Require one boundary condition for each inflow characteristic.

The question is, what do we specify?

The general form of the m-q BCs would be

$$\underline{\underline{B}}\;\underline{u}(0,t) = \underline{g}(t)$$

where
$$\underline{\underline{B}} = \underbrace{\begin{bmatrix} - & \underline{b}_1^T & - \\ & \vdots & \\ - & \underline{b}_{m-q}^T & - \end{bmatrix}}_{m} \Big\}\, m-q \quad , \quad \underline{g}(t) = \begin{bmatrix} \\ \\ \\ \end{bmatrix} \Big\}\, m-q$$

What are the row vectors $\underline{b}_1, \dots, \underline{b}_{m-q}$?

The characteristic form of the equation, $\underline{w}(x,t) = \underline{\underline{R}}^{-1}\underline{u}(x,t)$

$$\rightarrow \quad \underline{w}_t + \underline{\underline{\Lambda}}\,\underline{w}_x = 0 \quad , \quad \underline{\underline{A}}\,\underline{\underline{R}} = \underline{\underline{R}}\,\underline{\underline{\Lambda}}$$

The boundary conditions for $\underline{w}$:

$$\underline{\underline{B}}\,\underline{\underline{R}}\,\underline{w}(0,t) = \underline{g}(t)$$

let $\tilde{\underline{\underline{B}}} = \underline{\underline{B}}\,\underline{\underline{R}} = \begin{bmatrix} \text{---}\ \underline{b}_1^T\ \text{---} \\ \vdots \\ \text{---}\ \underline{b}_{m-q}^T\ \text{---} \end{bmatrix} \begin{bmatrix} | & & | \\ \underline{r}_1 & \cdots & \underline{r}_m \\ | & & | \end{bmatrix}$ $\quad (m-q, m) \times (m, m) = (m-q, m)$

$$\rightarrow \quad \tilde{\underline{\underline{B}}}_{ij} = \begin{bmatrix} \underline{b}_i^T \cdot \underline{r}_j \end{bmatrix} \quad , \quad i = 1, \dots, m-q, \quad j = 1, \dots, m$$

$$\Rightarrow \quad \tilde{\underline{\underline{B}}}\,\underline{w}(0,t) = \underline{g}(t) \Rightarrow \begin{bmatrix} & \tilde{B} & \end{bmatrix} \begin{bmatrix} \\ \underline{w}(0,t) \\ \\ \end{bmatrix} = \begin{bmatrix} \underline{g}(t) \end{bmatrix}$$

we need to partition the matrix:

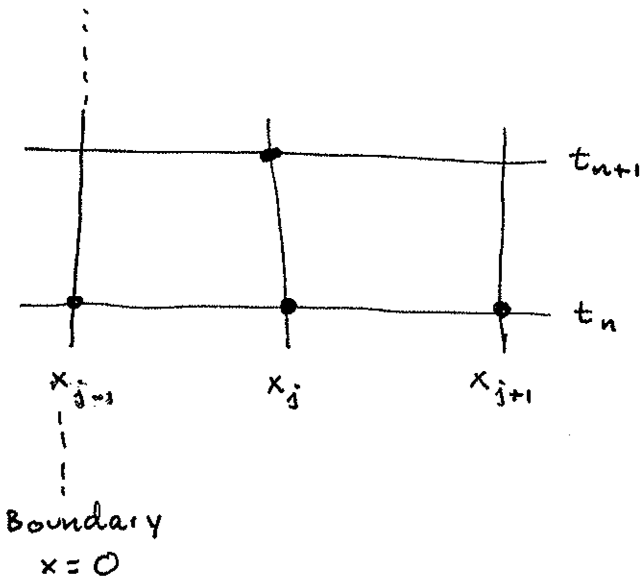
$$\begin{bmatrix} \tilde{\underline{\underline{B}}}_1 \,\vdots\, \tilde{\underline{\underline{B}}}_2 \end{bmatrix} \begin{bmatrix} \overbrace{w_1}^{\text{data from ICs}} \\ \cdots \\ \underbrace{w_2}_{\text{need to be specified by B.C.s}} \end{bmatrix} = \begin{bmatrix} \underline{g}(t) \end{bmatrix}$$

$$\tilde{\underline{\underline{B}}}_2\,\underline{w}_2(0,t) = \underline{g}(t) - \tilde{\underline{\underline{B}}}_1\,\underline{w}_1(0,t)$$

therefore we need that $\tilde{\underline{\underline{B}}}_2$ be nonsingular.

From a numerical point of view:

Many schemes, such as Lax-Friedrichs and Lax-Wendroff have the following stencils



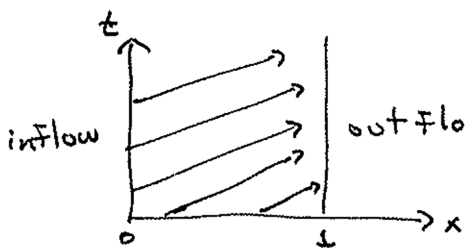
At $x=0$, require m-components for v_{j-1}^n, but only $m-q$ corresponding to $m-q$ inflow characteristics would be specified by the boundary conditions.

To complete the remaining q components of $\underline{v}_{j-1}^n$, use extrapolation from the interior.

Example: $u_t + u_x = 0$, $0 < x < 1$, $t > 0$

$u(x,0) = f(x)$

$u(0,t) = g(t)$ (inflow boundary)



numerical approximation:

$x_j = j\Delta x$, $\Delta x = \frac{1}{N}$, $t_n = n\Delta t$ $\quad j = 1, \ldots, N-1$

$$v_j^{n+1} = \frac{1}{2}\left(v_{j-1}^n + v_{j+1}^n\right) - \frac{\Delta t}{\Delta x}\left(v_{j+1}^n - v_{j-1}^n\right)$$ - Lax-Friedrichs

$v_j^0 = f(x_j)$, $0 \le j \le N$

$v_0^n = g(t_n)$, $n > 0$

v_N^n = extrapolation, simplest is $v_N^n = v_{N-1}^n$

Hyperbolic Conservation Laws

Basic equation: $\boxed{\underline{u}_t + \underline{f}(\underline{u})_x = 0}$ conservation form

where $\underline{u}(x,t) = m$ "state" variables

$\underline{f}(\underline{u}(x,t)) = m$ "flux" functions

This general form is derived from an integral conservation:

integrate differential form from $x=a$ to $x=b$

$$\int_a^b \left(u_t + \underline{f}(\underline{u})_x\right) dx = 0$$

$$\rightarrow \quad \int_a^b u_t \, dx + \underline{f}(\underline{u}(x,t)) \Big]_{x=a}^{x=b} = 0$$

$$\rightarrow \quad \boxed{\frac{d}{dt} \int_a^b \underline{u}(x,t)\, dx + \underline{f}(\underline{u}(x,t)) \Big]_{x=a}^{x=b} = 0} \quad \text{integral form}$$

The rate of change of $\underline{u}$ between $x=a$ and $x=b$ is balanced by the net flux of $\underline{u}$ at the boundaries

If $\left[\underline{f}(\underline{u})\right]_a^b = 0$ then $\int_a^b \underline{u}(x,t)\,dx = \text{constant}$

and $\underline{u}$ is a vector of conserved quantities.

$$\boxed{\underline{u}_t + \underline{f}_{\underline{u}}(\underline{u})\, \underline{u}_x = 0}, \quad \text{quasilinear form}$$

where $\underline{f}_{\underline{u}}$ is a Jacobian matrix. If this Jacobian matrix is diagonalizable with real eigenvalues for any $\underline{u}$ then this equation is hyperbolic.

Examples

- scalar equations
  - let $f = cu$, $c = \text{const}$ $\rightarrow$ $u_t + (cu)_x = 0$, conservation form

    $u_t + cu_x = 0$, quasilinear form

  - let $f = \frac{1}{2}u^2$ (nonlinear) Burgers Inviscid Equation

    $u_t + \left(\frac{1}{2}u^2\right)_x = 0$, conservation form

    $u_t + uu_x = 0$, quasilinear form

- systems
  - $\underline{F} = \underline{\underline{A}}\,\underline{u}$, $\underline{\underline{A}}$ real eigenvalues

    $\underline{u}_t + (\underline{\underline{A}}\,\underline{u}_x) = 0$, conservation form

    $\underline{u}_t + \underline{\underline{A}}\,\underline{u}_x = 0$, quasilinear form

  - $\underline{F} = \begin{bmatrix} u_2 \\ u_2^2/u_1 + a^2 u_1 \end{bmatrix}$, $a > 0$, constant, $\underline{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

u_1 is a density $\Rightarrow$ $u_1 > 0$

u_2 is a momentum

a is a sound speed

$\underline{u}_t + \underline{F}(\underline{u})_x = 0$

$\underline{u}_t + \underline{F}_{\underline{u}}(\underline{u})\,\underline{u}_x = 0$

calculate Jacobian

$$\underline{F}_{\underline{u}} = \begin{bmatrix} \frac{\partial F_{11}}{\partial u_1} & \frac{\partial F_{12}}{\partial u_2} \\ \frac{\partial F_{21}}{\partial u_1} & \frac{\partial F_{22}}{\partial u_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{u_2^2}{u_1^2} + a^2 & \frac{2u_2}{u_1} \end{bmatrix}$$

$\boxed{\text{Acoustic waves}}$

$$\det \begin{bmatrix} -\lambda & 1 \\ a^2 - \frac{u_2^2}{u_1^2} & \frac{2u_2}{u_1} - \lambda \end{bmatrix} = \lambda\left(\lambda - \frac{2u_2}{u_1}\right) - \left(a^2 - \frac{u_2^2}{u_1^2}\right)$$

$$= \lambda^2 - \frac{2u_2}{u_1}\lambda - \left(a^2 - \frac{u_2^2}{u_1^2}\right) = 0$$

$\rightarrow$ $\boxed{\lambda = \frac{u_2}{u_1} \pm a}$, by quadratic equation

$\Rightarrow$ $\lambda = v \pm a$

Scalar Conservation Laws

Begin with the scalar case: (pure IVP)

$$u_t + f(u)_x = 0, \quad |x| < \infty, \quad t > 0, \quad u(x,0) = u_0(x)$$

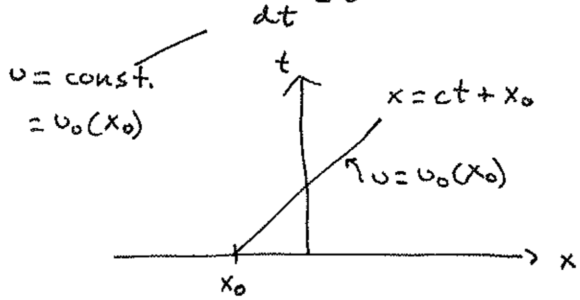
Consider case $f(u) = cu$, $c = \text{const.}$

$$\rightarrow u_t + c u_x = 0, \quad u(x,0) = u_0(x)$$

characteristic description:

$$\frac{du}{dt} = 0 \quad \text{along} \quad \frac{dx}{dt} = c$$

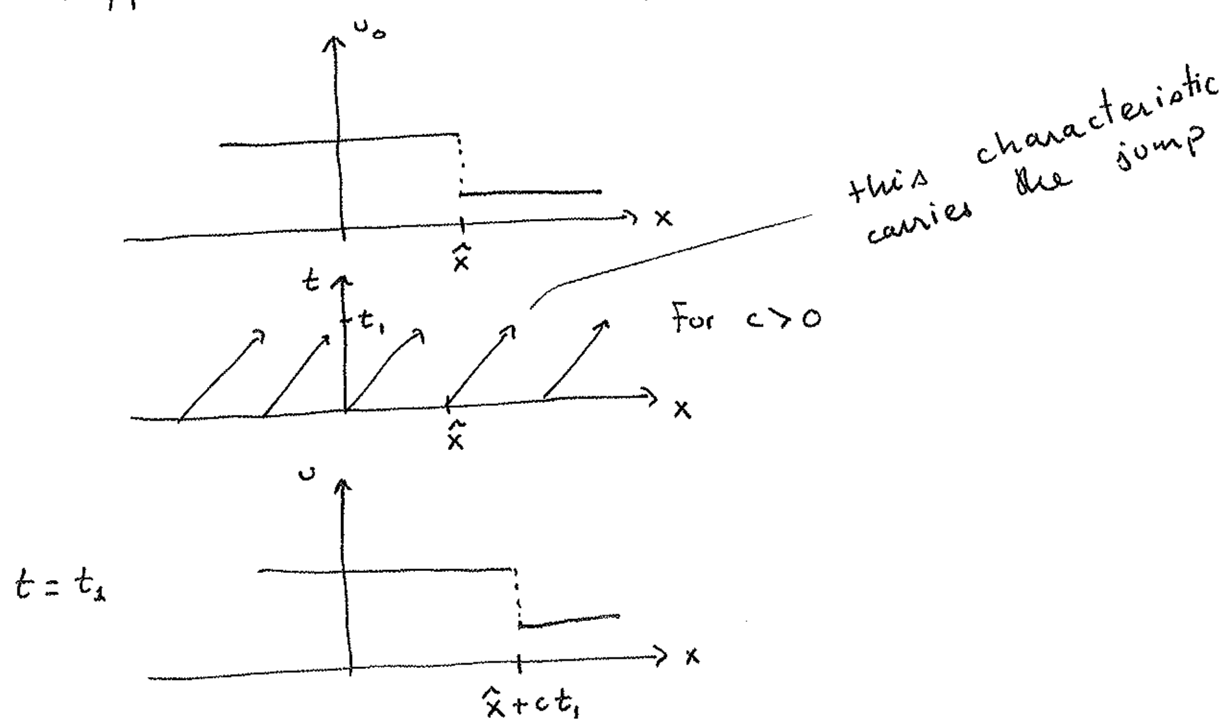
$\hookrightarrow \; x = ct + x_0$, where $x_0 = \text{const}$



$$\boxed{\begin{array}{l} u(x,t) = u_0(x - ct) \\ \text{shape of initial} \\ \text{function translates} \\ \text{with speed } c \end{array}}$$

Of particular interest are solutions with discontinuities

Suppose $u_0(x)$ has a jump discontinuity at $\hat{x}$.



Need to interpret discontinuities in the PDE.

1) zero limit of a "viscous" PDE

$$\tilde{u}_t + c\tilde{u}_x = \nu \tilde{u}_{xx} \quad , \quad \nu = \text{const} > 0$$

$$\tilde{u}(x,0) = u_0(x)$$

interpret solution of $u_t + cu_x = 0$, $u(x,0) = u_0(x)$, the inviscid problem as the limit as $\nu \to 0$.

change of variables: let $\xi = x - ct$, $\hat{t} = t$

$$\Rightarrow \quad \tilde{u}_{\hat{t}} = \nu \tilde{u}_{\xi\xi} \quad , \quad \tilde{u}(\xi,0) = u_0(\xi) \qquad \text{heat equation}$$

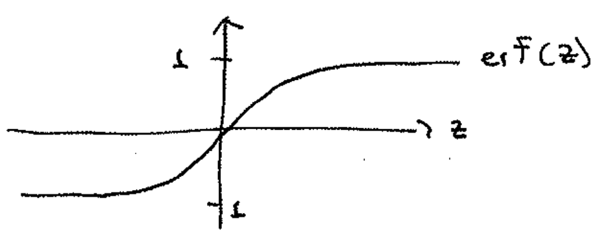
solve using Fourier transforms

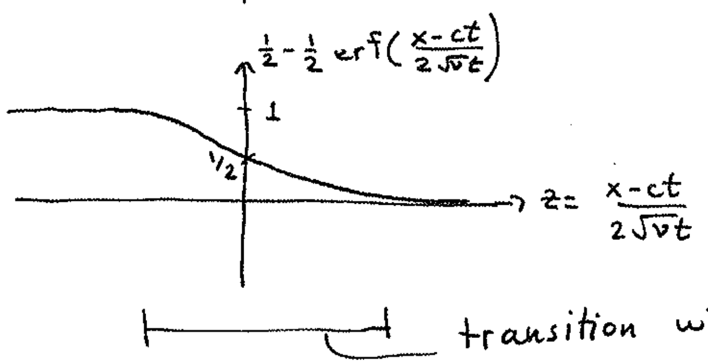
$$\hat{u}(\xi,t) = \frac{1}{\sqrt{4\pi\nu t}} \int_{-\infty}^{\infty} u_0(s)\, e^{-(s-\xi)^2/4\nu t}\, ds$$

If $u_0(x) = \begin{cases} 1 & , x<0 \\ 0 & , x>0 \end{cases} \quad \rightarrow \quad \hat{u}(\xi,t) = \frac{1}{\sqrt{4\pi\nu t}} \int_{-\infty}^{0} e^{-(s-\xi)^2/4\nu t}\, ds$

$$\rightarrow \quad \hat{u}(\xi,t) = \frac{1}{2} - \frac{1}{2}\,\text{erf}\left(\frac{\xi}{2\sqrt{\nu t}}\right)$$

$$\rightarrow \quad \tilde{u}(x,t) = \frac{1}{2} - \frac{1}{2}\,\text{erf}\left(\frac{x-ct}{2\sqrt{\nu t}}\right) \quad \text{where} \quad \boxed{\text{erf}\, z = \frac{2}{\sqrt{\pi}} \int_0^z e^{-s^2}\, ds}$$





transition width decreases as ν gets ~~large~~ smaller and in the limit as $\nu \to 0$, becomes a jump.

2) Weak solutions of the integral conservation

$$\frac{d}{dt}\int_a^b u\,dx + f(u)\Big|_a^b = 0$$

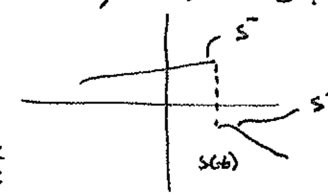
suppose $u(x,t)$ has a jump discontinuity at $x = s(t)$

Position a, b on either side of $x = s(t)$.

$$\rightarrow \frac{d}{dt}\int_a^b u\,dx = \frac{d}{dt}\int_a^{s(t)} u\,dx + \frac{d}{dt}\int_{s(t)}^b u\,dx$$

then u is smooth for each interval $a < x < s(t)$, $s(t) < x < b$.

so move derivative inside



$$\rightarrow \frac{d}{dt}\int_a^b u\,dx = \int_a^{s(t)} u_t\,dx + \int_{s(t)}^b u_t\,dx + u(s^-,t)\frac{ds}{dt} - u(s^+,t)\frac{ds}{dt}$$

substitute $u_t = -f_x(u)$

$$\rightarrow \frac{d}{dt}\int_a^b u\,dx = -\int_a^{s(t)} f_x(u)\,dx - \int_{s(t)}^b f_x(u)\,dx + \left(u(s^-,t) - u(s^+,t)\right)\frac{ds}{dt}$$

$$= -f(u)\Big]_a^{s(t)} - f(u)\Big]_{s(t)}^b + \left(u(s^-,t) - u(s^+,t)\right)\frac{ds}{dt}$$

$$\frac{d}{dt}\int_a^b u\,dx = -f(u(s^-,t)) + f(u(a,t)) + f(u(s^+,t)) - f(u(b,t)) + \left(u(s^-,t) - u(s^+,t)\right)\frac{ds}{dt}$$

Define jump notation:

$$[u] \equiv u(s^+,t) - u(s^-,t)$$

$$[f] \equiv f(u(s^+,t)) - f(u(s^-,t))$$

$$\rightarrow \frac{d}{dt}\int_a^b u\,dx = [f] - f(u)\Big|_a^b - [u]\frac{ds}{dt}$$

recall, $\frac{d}{dt}\int_a^b u\,dx + f(u)\Big|_a^b = 0$

$$\Rightarrow \boxed{[u]\frac{ds}{dt} = [f]}$$ Rankine-Hugoniot jump conditions

Let us apply this to the linear case,

$$u_t + cu_x = 0$$

where $f = cu$

$$\rightarrow [u]\frac{ds}{dt} = [cu] \rightarrow [u]\frac{ds}{dt} = c[u] \rightarrow \frac{ds}{dt} = c$$

weak form: use PDE whenever solution is smooth and patch u across any jump using jump conditions

Burgers Equation

(simplest nonlinearity - quadratic)

$f(u) = \frac{1}{2}u^2$

$u_t + (\frac{1}{2}u^2)_x = 0$, $u(x,0) = u_0(x)$ where u_0 is smooth

$$u_t + f(u)_x = 0$$

$$u_t + f'(u)\, u_x = 0 \quad , \quad c(u) = f'(u)$$

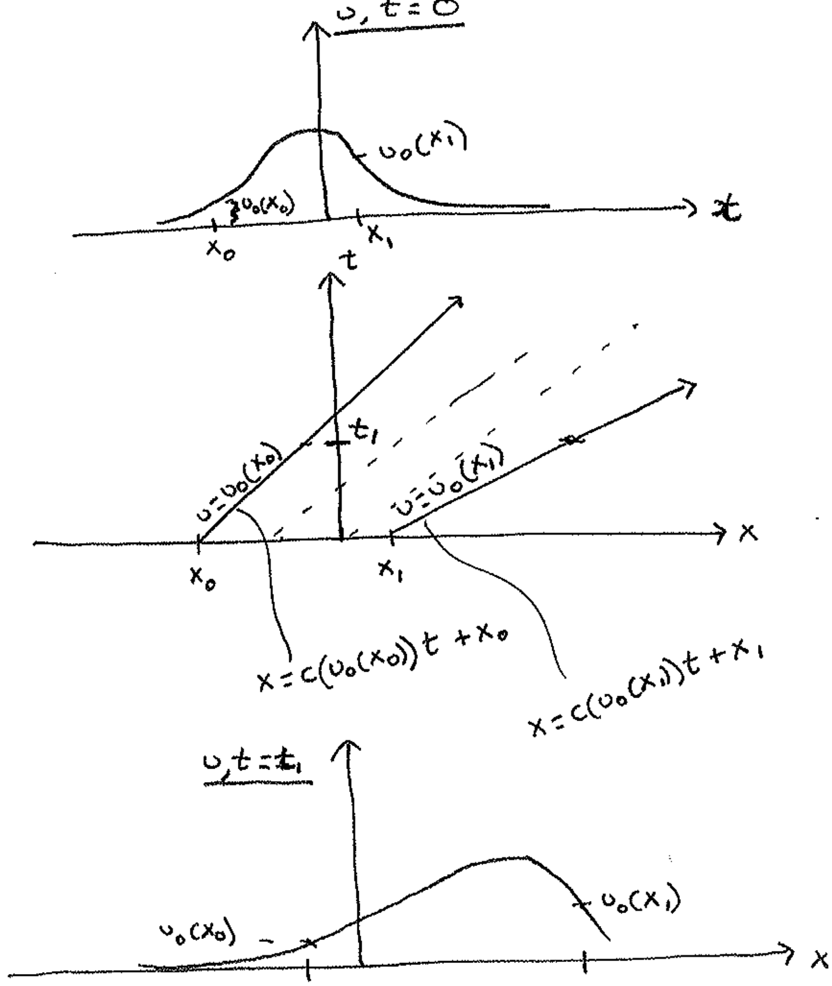
<u>characteristics:</u> $\dfrac{du}{dt} = 0$ along $\dfrac{dx}{dt} = c(u)$

$\rightarrow$ $u = \text{const.}$ along a characteristic

$\rightarrow$ $c(u)$ is a constant along a characteristic

$\rightarrow$ $x = c(u)\, t + \text{const.}$

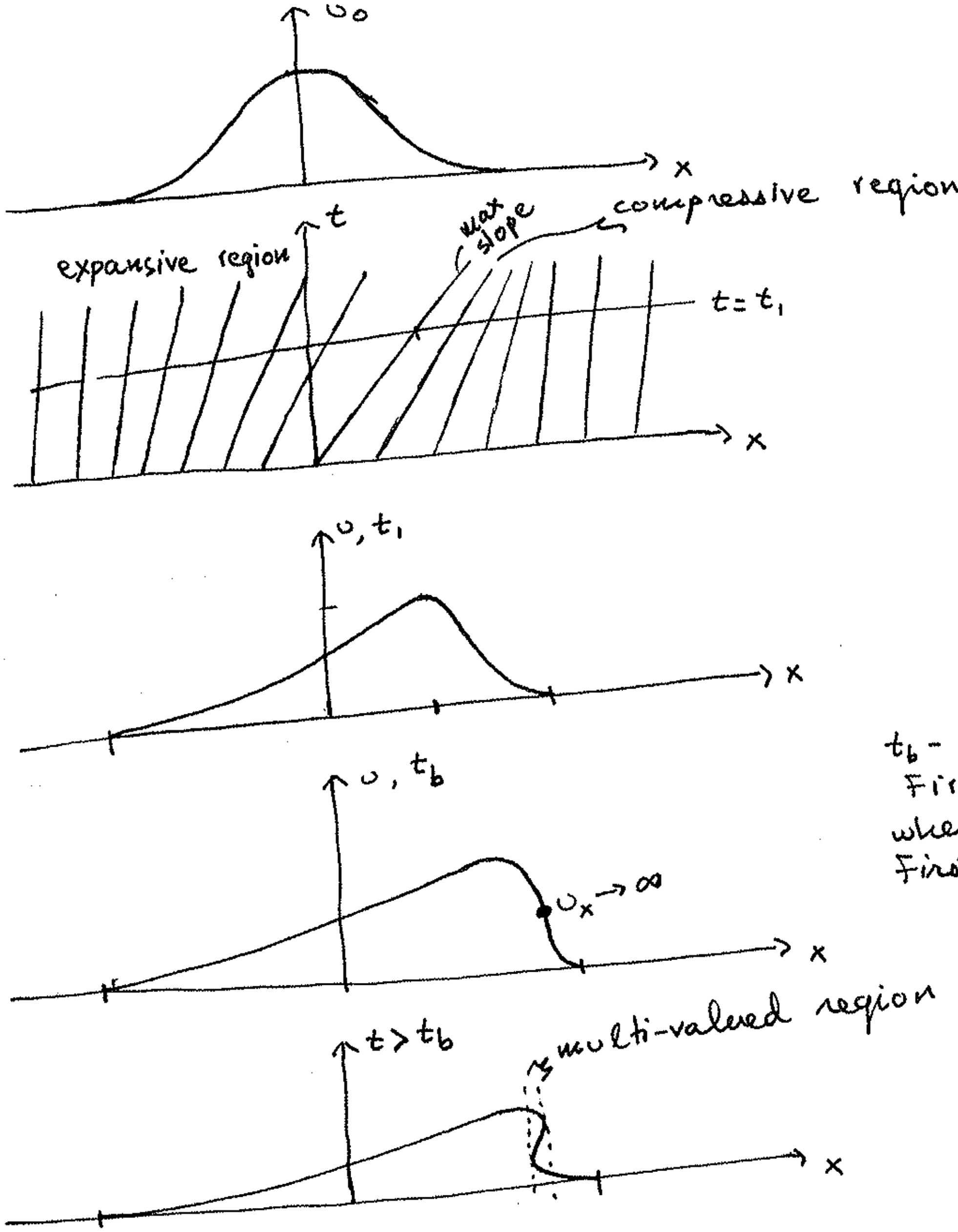
<u>Graphical construction:</u>



initial shape given by $u_0(x)$ translates and distorts

Burger's eqn: $f(u) = \frac{1}{2}u^2$

$c(u) = f'(u) = u$



t_b - time of first breaking, when characteristics first cross

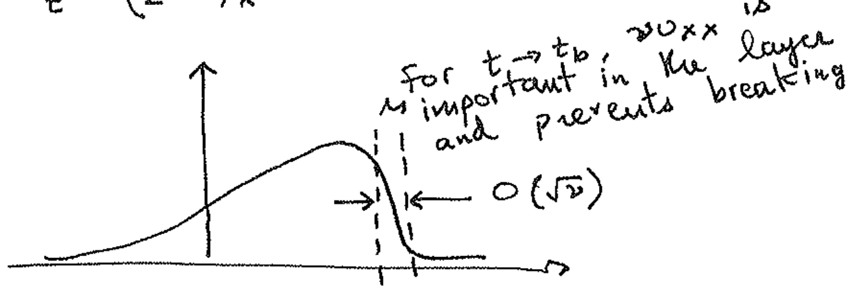
this solution satisfies the PDE for $t < t_b$

At $t = t_b$, a singularity in the PDE occurs, so this solution is no longer acceptable.

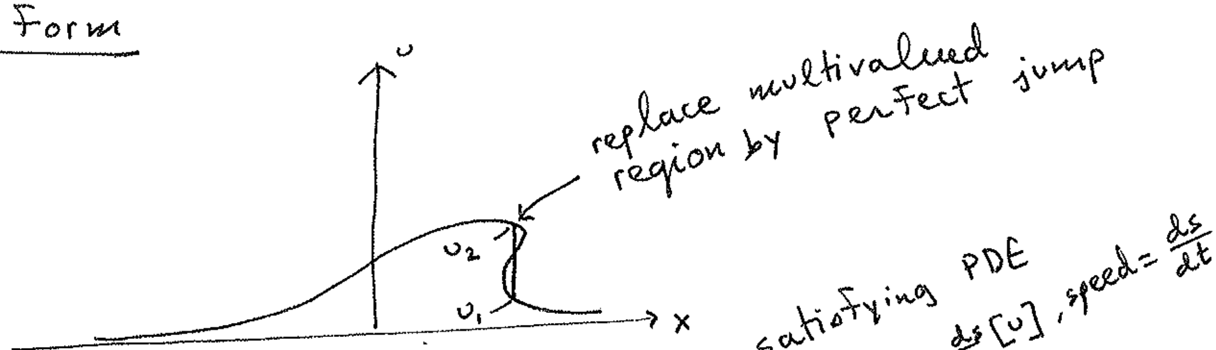
we need to recover the solution

1) <u>viscous limit,</u> $u_t + \left(\frac{1}{2}u^2\right)_x = \nu u_{xx}$, $\nu \to 0$

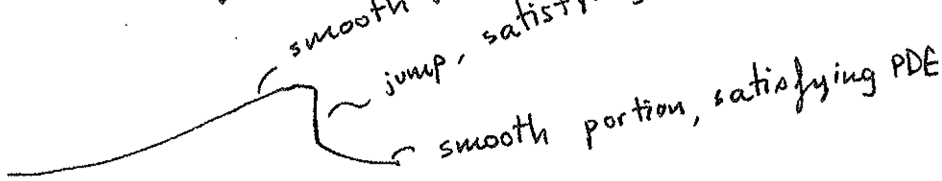
<u>near $t = t_b$</u> ;



2) <u>weak form</u>



$\Downarrow$



For Burgers eqn:

$$[f] = \frac{ds}{dt}[u]$$

$$\frac{1}{2}\left(u_2^2 - u_1^2\right) = \frac{ds}{dt}(u_2 - u_1)$$

the solution is constructed by patching smooth bits where the PDE holds together with jumps where the jump condition holds.

<u>The issue with the weak solution is that it is NOT UNIQUE.</u>

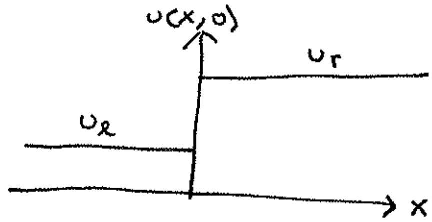
We can force uniqueness by insisting that the weak form is the $\nu \to 0$ limit of the viscous problem

$\to$ entropy satisfying solutions

Example: "Riemann" problem

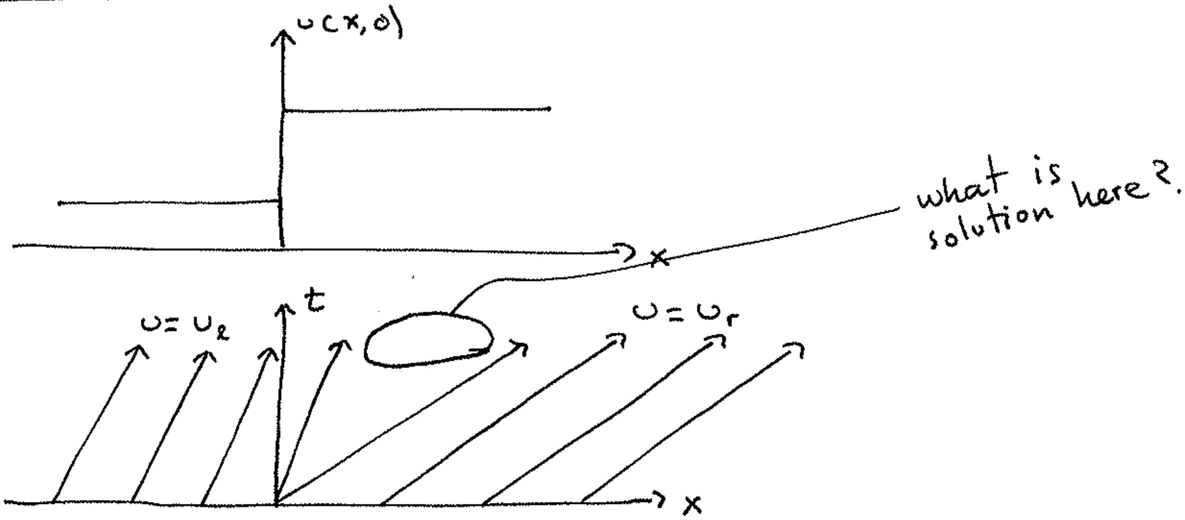
$$u_t + \left(\tfrac{1}{2}u^2\right)_x = 0 \quad , \quad |x| < \infty \, , \quad t > 0$$

$$u(x,0) = \begin{cases} u_l \, , & x < 0 \\ u_r \, , & x > 0 \end{cases}$$



- there are 2 cases $u_l < u_r$ and $u_l > u_r$

case (1) $u_l < u_r$



similarity form: let $u(x,t) = w(\eta)$, $\eta = \frac{x}{t}$

$$\rightarrow \; u_t = w' \cdot \eta_t = -w' \frac{x}{t^2} = -\eta w'(\eta) \frac{1}{t}$$

$$u_x = w'(\eta)\, \eta_x = w'(\eta) \cdot \frac{1}{t}$$

$$\rightarrow \; u_t + u \cdot u_x = 0 \;\rightarrow\; -\eta\, w'(\eta) \cdot \frac{1}{t} + w \cdot w'(\eta) \frac{1}{t} = 0$$

$$\rightarrow \; (w - \eta)\, w'(\eta) = 0$$

$$\rightarrow \; w'(\eta) = 0 \;\rightarrow\; w(\eta) = \text{const} \quad \text{or} \quad w = \eta$$

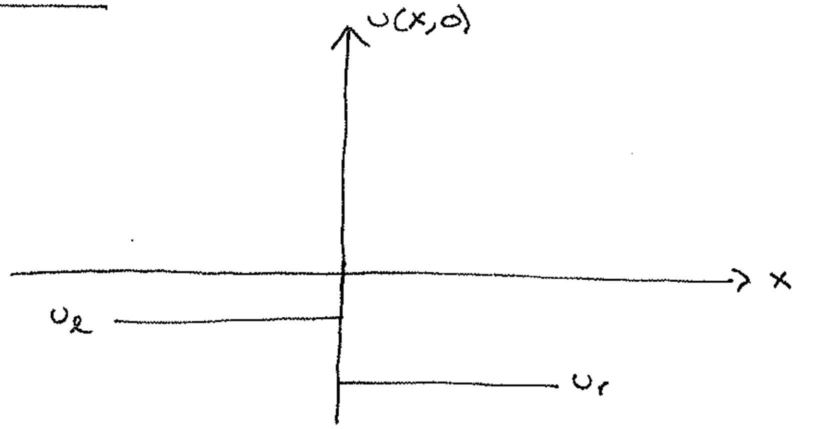
$$\downarrow \qquad\qquad \downarrow$$

$$u = \text{const} \qquad\qquad u = \frac{x}{t}$$

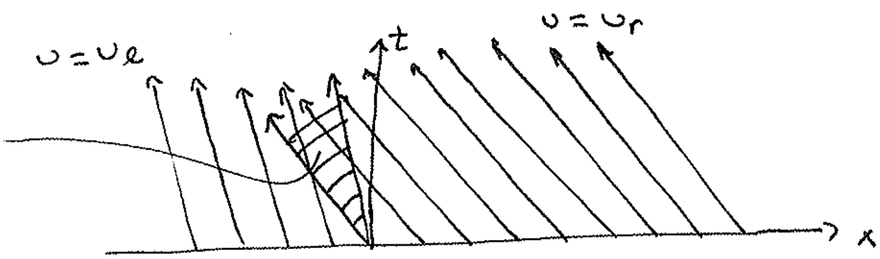
in solution, notice that $u = \text{const}$ outside expansive region and inside, we apply $u = \frac{x}{t}$

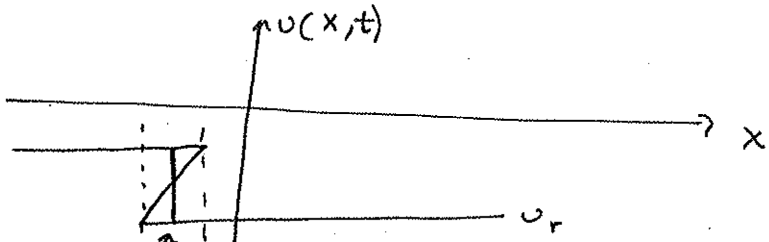
$$\boxed{\text{the solution is} \quad u(x,t) = \begin{cases} u_\ell \ , & \frac{x}{t} < u_\ell \\ \frac{x}{t} \ , & u_\ell < \frac{x}{t} < u_r \\ u_r \ , & \frac{x}{t} > u_r \end{cases}}$$

<u>Case ②</u> $u_\ell > u_r$



region of overlap





replace multivalued region by a jump that propagates at speed $\frac{ds}{dt}$,

$[f] = \frac{ds}{dt}[u]$, $s(t)$ - position of jump

For Burger's equation

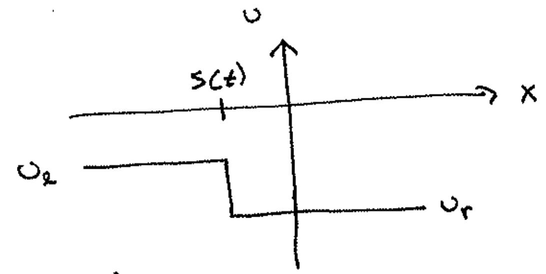
$$[f(u)] = [u] \frac{ds}{dt}$$

$$\frac{1}{2}(u_r^2 - u_\ell^2) = (u_r - u_\ell) \frac{ds}{dt}$$

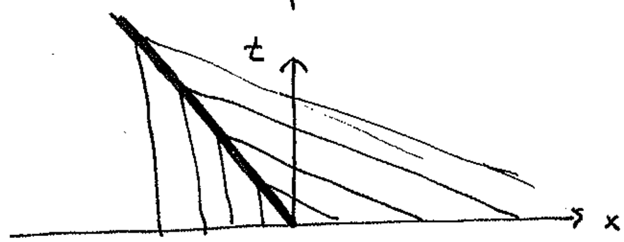
$$\rightarrow \boxed{\frac{ds}{dt} = \frac{u_r + u_\ell}{2}} \rightarrow s(t) = \left(\frac{u_r + u_\ell}{2}\right) t + \underset{0}{\text{const}}$$

$$\downarrow$$

$$\boxed{s(t) = \left(\frac{u_r + u_\ell}{2}\right) t}$$

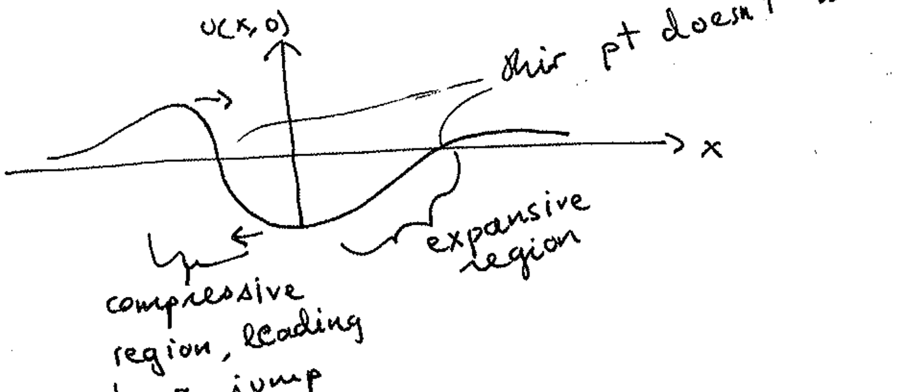


$$s(t) = \left(\frac{u_r + u_l}{2}\right) t$$



$$\boxed{u(x,t) = \begin{cases} u_l, & x < \frac{1}{2}(u_l + u_r)t \\ u_r, & x > \frac{1}{2}(u_l+u_r)t \end{cases}}$$

For example, $u_t + \left(\frac{1}{2}u^2\right)_x = 0$



Numerics (for scalar case)

$u_t + f(u)_x = 0, \quad |x| < \infty, \quad t > 0$

$u(x,0) = u_0(x)$

Issues (1) If solution is smooth, then the nonlinearity does not play a significant role and a numerical approximation should behave as if the equation is linear.

(2) Nonlinearity can lead to jump discontinuities forming in finite time. We need to worry about how numerical approximation behaves near jumps and what equation is being approximated in the vicinity of a jump.

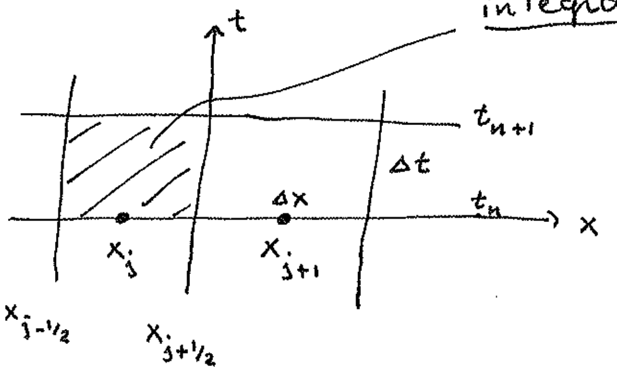
3) Uniqueness of weak solutions and nonlinear stability.

Finite Volume Formulation

$\Rightarrow$ try to maintain the integral formulation (integral conservation) in the discrete approximation

$\Rightarrow$ to obtain correct weak form

integrate over a cell



$$u_t + f(u)_x = 0$$

$$\int_{t_n}^{t_{n+1}} \int_{x_{j-1/2}}^{x_{j+1/2}} \left(u_t + f(u)_x\right) dx\, dt = 0$$

$$\rightarrow \quad \left[\int_{x_{j-1/2}}^{x_{j+1/2}} u\, dx\right]_{t_n}^{t_{n+1}} + \left[\int_{t_n}^{t_{n+1}} f(u)\, dt\right]_{x_{j-1/2}}^{x_{j+1/2}} = 0$$

everything is exact

Define $$U_j^n \equiv \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t_n)\, dx$$ "cell average"

$$F_{j+1/2}^n \equiv \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} f\left(u(x_{j+1/2}, t)\right) dt$$

$$\rightarrow \quad \Delta x\, U_j^{n+1} - \Delta x\, U_j^n + \Delta t\, F_{j+1/2}^n - \Delta t\, F_{j-1/2}^n = 0$$

$$\rightarrow \quad \boxed{U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x}\left(F_{j+1/2}^n - F_{j-1/2}^n\right) \quad \text{still exact}}$$

Numerical approximation comes from the approximations of $F^n_{j+1/2}$ and $F^n_{j-1/2}$ ie numerical flux functions

typically, $F^n_{j+1/2} \simeq \underset{\substack{\uparrow \\ \text{numerical flux function}}}{G}(u^n_j, u^n_{j+1})$, $F^n_{j-1/2} \simeq G(v^n_{j-1}, v^n_j)$

comment, there are other possibilities to define F, eg: $F^n_{j+1/2} = \tilde{G}(v^n_{j-p}, \dots, v^n_{j+q})$

Conservative Finite Volume Scheme

$$v^{n+1}_j = v^n_j - \frac{\Delta t}{\Delta x}\Big(G(v^n_j, v^n_{j+1}) - G(v^n_{j-1}, v^n_j)\Big)$$

where $\displaystyle v^n_j \simeq \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x, t_n)\, dx$

and $\displaystyle G(v^n_j, v^n_{j+1}) \simeq \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} f(u(x_{j+1/2}, t))\, dt$

note, $\displaystyle v^0_j = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\, dx = u_0(x_j) + O(\Delta x^2)$

Examples

$$\cancel{G(u_\ell} \quad G(u_\ell, u_r) = \frac{1}{2}\Big(f(u_\ell) + f(u_r)\Big) - \frac{\Delta x}{2\Delta t}\Big(u_r - u_\ell \Big)$$

Lax-Friedrichs (1st order)

$$G(u_\ell, u_r) = \frac{1}{2}\Big(f(u_\ell) + f(u_r)\Big) - \frac{\Delta t}{2\Delta x} f'(\bar{a})\Big(f(u_r) - f(u_\ell)\Big)$$

Lax-Wendroff (2nd order)

where $\bar{a} = \frac{1}{2}(u_\ell + u_r)$

11/03/03

Comments

- There are many more choices for G, other than Lax-Friedrichs and Lax-Wendroff.
- Not free to make any choice for $G(u_\ell, u_r)$.

Consistency requires that

$$G(u,u) = f(u)$$

and smoothness of function G.

Consider the truncation error: — consider function with two arguments $G(u,v)$

$$\tau_j^n = \frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{G(u_j^n, u_{j+1}^n) - G(u_{j-1}^n, u_j^n)}{\Delta x}$$

$$\tau_j^n = u_t + O(\Delta t) + \frac{G(u, u + \Delta x u_x + \dots) - G(u - \Delta x u_x + \dots, u)}{\Delta x}$$

$$\tau_j^n = u_t + O(\Delta t) + \frac{\cancel{G(u,u)} + G_v(u,u)\Delta x u_x + O(\Delta x^2)}{\Delta x} - \frac{\cancel{G(u,u)} - G_u(u,u)\Delta x u_x + O(\Delta x^2)}{\Delta x}$$

$$\tau_j^n = u_t + O(\Delta t) + \left(G_u(u,u) + G_v(u,u)\right) u_x + O(\Delta x)$$

we want $f'(u) = \left(G_u(u,u) + G_v(u,u)\right)$ for consistency

this is true if $f(u) = G(u,u)$.

- The method is conservative, which means that if jump discontinuities develop $\to$ then they will propagate with the correct speed as determined by the integral conservation $\Rightarrow$ $[f(u)] = [u]\frac{ds}{dt}$

Example

$$u_t + f(u)_x = 0, \quad |x| < \infty, \quad t > 0, \quad u(x,0) = u_0(x)$$

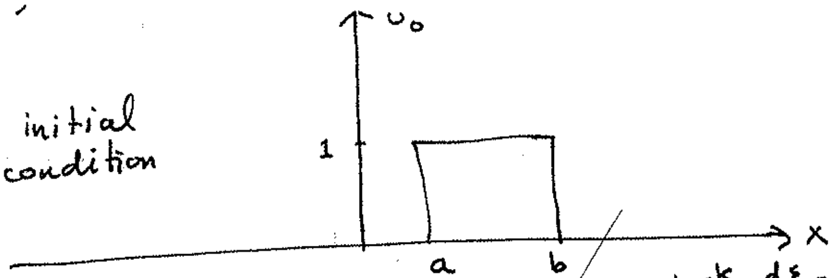
Solve using

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left(G(v_j^n, v_{j+1}^n) - G(v_{j-1}^n, v_j^n) \right)$$

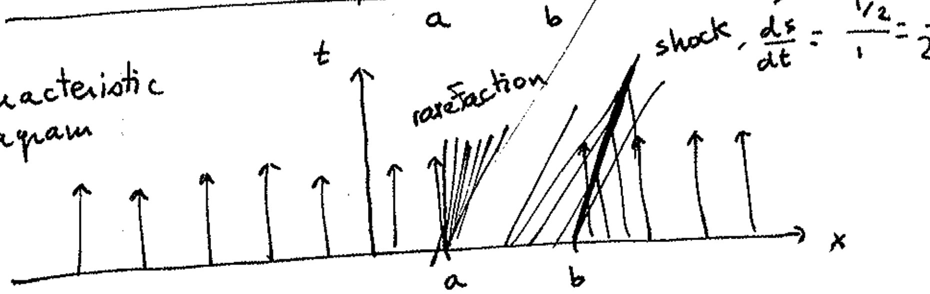
using both Lax-Friedrichs and Lax-Wendroff for G.

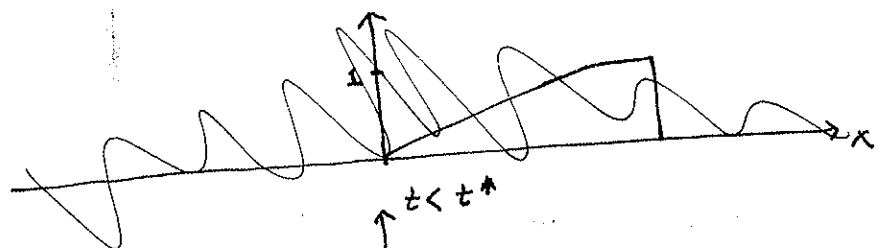
Assume $f(u) = \frac{1}{2}u^2$, inviscid Burgers eqn, and that initial state is

initial condition

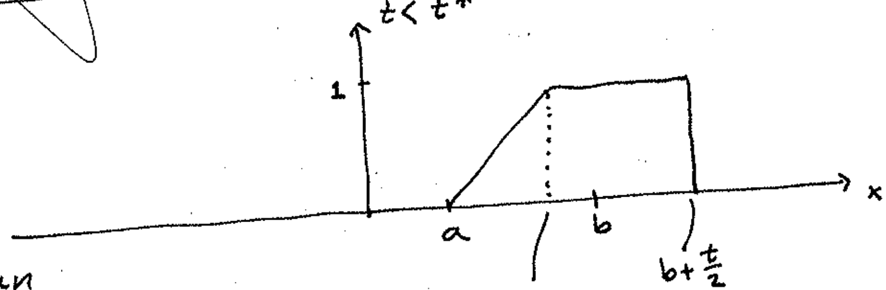


characteristic diagram



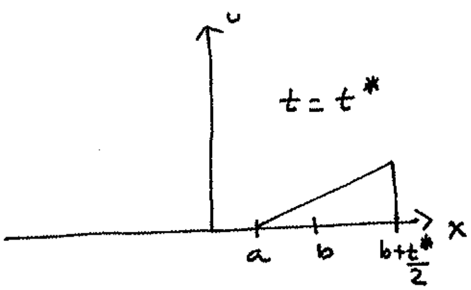


$$\Rightarrow \boxed{u(x,t) = \begin{cases} 0, & x < a \\ \frac{x-a}{t}, & a < x < a+t \\ 1, & a+t < x < b + \frac{t}{2} \\ 0, & x > b + \frac{t}{2} \end{cases}}$$

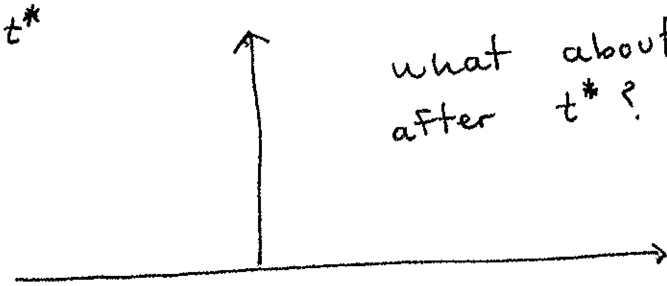


Notice, the expansion fan meets the shock when

$$a + t = b + \frac{t}{2} \quad \rightarrow \quad t = 2(b-a) = t^*$$



what about after t^*? $\Rightarrow$



After t^*, $\frac{ds}{dt} = \frac{f(\hat{u}) - f(0)}{\hat{u} - 0}$

where $\hat{u}$ is the solution in the expansion fan at $x = s(t)$, ie at the shock

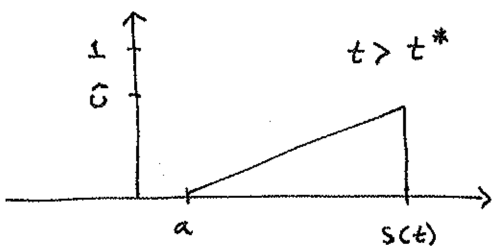
$f(0) = 0$, $\rightsquigarrow$ $\frac{ds}{dt} = \frac{\frac{1}{2}\hat{u}^2}{\hat{u}} = \frac{1}{2}\hat{u} = \frac{1}{2}\left(\frac{s-a}{t}\right)$

solve differential equation for s, with initial condition

$s(t^*) = b + \frac{t^*}{2}$

$\Rightarrow$ $s(t) = \sqrt{2t(b-a)} + a$, $t \geq t^*$

$\hat{u} = \sqrt{\frac{2(b-a)}{t}}$



Test the numerical scheme using following parameters:

$x_j = (j - 1/2)\Delta x$, $\Delta x = \frac{1}{N}$

$v_j^0 = \frac{1}{2}\tanh(\lambda(x_j - a)) + \frac{1}{2}\tanh(\lambda(b - x_j))$

$a = 0.1$, $b = 0.4$, $\lambda = 100$, $N = 200$

Compare with non-conservative scheme

$u_t + u u_x = 0$, for $u \geq 0$

$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x} v_j^n \left(v_j^n - v_{j-1}^n\right)$

results in completely wrong shock location

Lax-Wendroff Theorem (LeVeque)

Consider a sequence of grids indexed by $k = 1, 2, \ldots$ with Δx_k and Δt_k vanishing as $k \to \infty$. Let $v_k(x,t)$ be a piecewise constant function taking the value v_j^n when $x \in (x_{j-1/2}, x_{j+1/2}]$, $t \in [t_n, t_{n+1})$ on grid k, where v_j^n is obtained from a consistent, conservative scheme. If v_k converges to a function $u(x,t)$ as $k \to \infty$, then u is a weak solution of the conservation law.

Remember, weak solutions aren't necessarily unique!

Remarks

1) The Lax-Wendroff theorem does not guarantee convergence. It applies if the scheme converges. If (Require nonlinear stability analysis. If solution is smooth, stability analysis to linearized solution applies, but when jumps occurr, this no longer applies.)

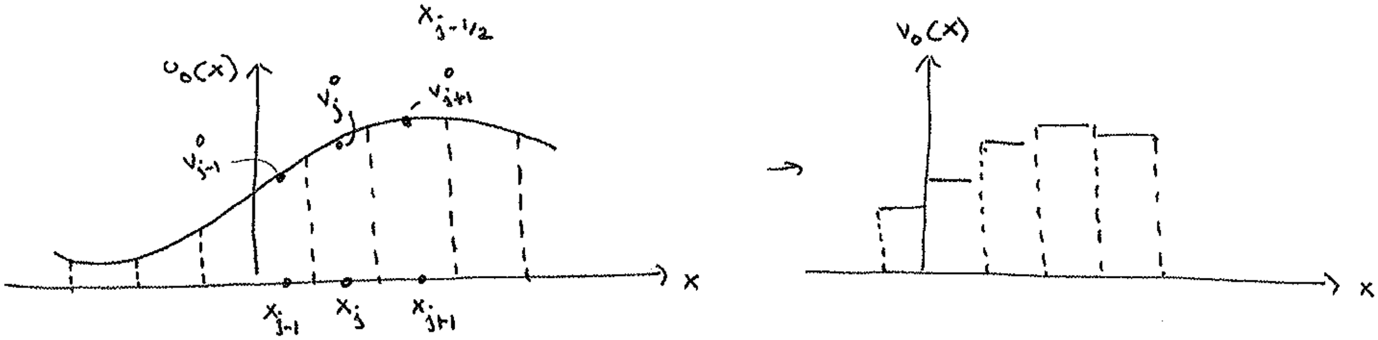
2) If convergence occurs, then discrete solution converges to weak solution, but the weak solutions are not unique! Require solutions that are entropy satisfying...

Godunov Methods

Essentially a nonlinear version of upwind methods. Numerical approach is still conservative and based on solutions of Riemann problems.

$$u_t + f(u)_x = 0 \quad , \quad u(x,0) = u_0(x)$$

Define $\displaystyle v_j^0 = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\,dx$ = cell average of initial data

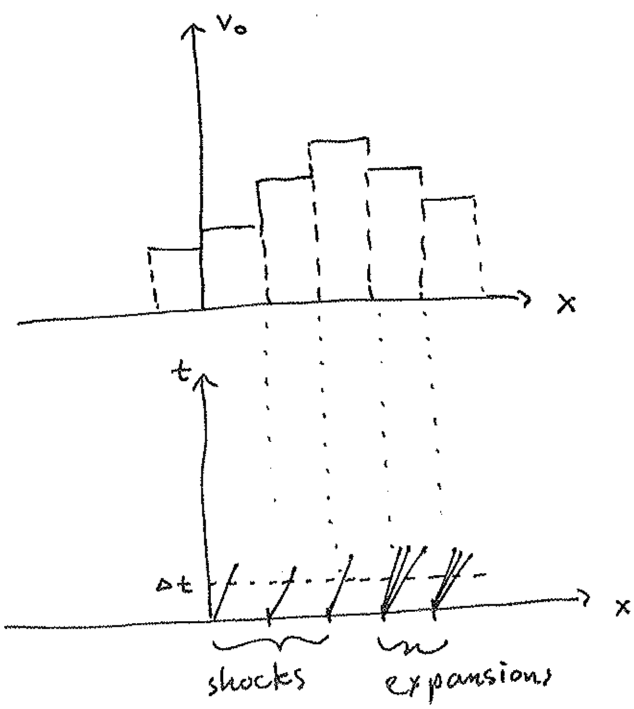


Let $v_0(x)$ be piecewise constant, $= v_j^0$ for $x \in x_{j-1/2}, x_{j+1/2}$

Consider the exact solution of the problem

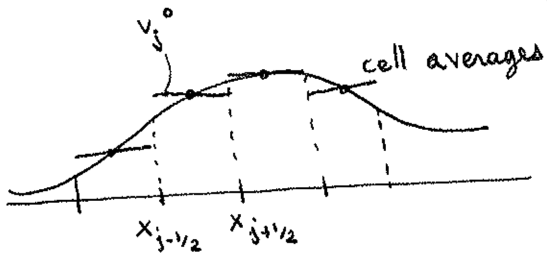
$$u_t + f(u)_x = 0$$

$u(x,0) = v_0(x)$ – piecewise constant approximation of u_0



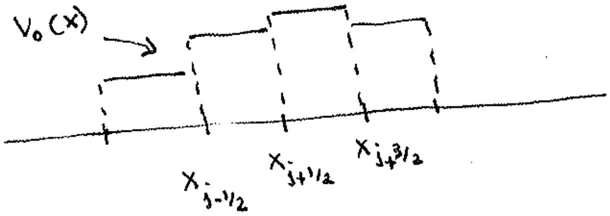
Godunov - reconstruct solution at Δt exactly based on $u(x,0) = v_0(x)$

cell average: $v_j^0 = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} v_0(x)\,dx$



From initial cell averages, construct a piecewise constant function $v_0(x)$ st

$v_0(x) = v_j^0$ for $x \in (x_{j-1/2}, x_{j+1/2})$

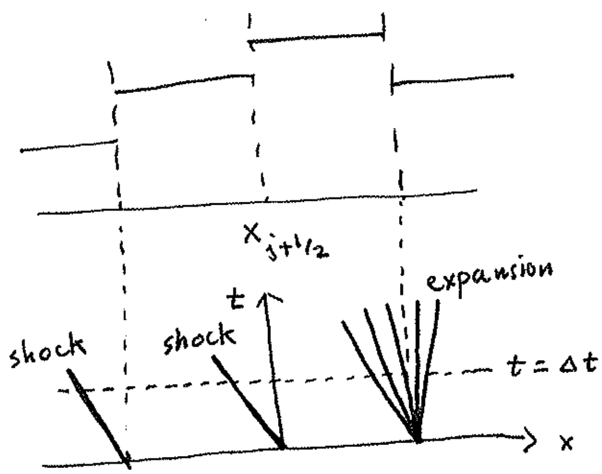


$\leftarrow$ solve the IVP:

$$\boxed{\begin{aligned} &u_t + f(u)_x = 0, \quad |x| < \infty, \quad 0 < t < \Delta t \\ &u(x,0) = v_0(x) \end{aligned}}$$

so instead of solving problem where initial data is smooth, solve same problem where initial data is piecewise constant $\rightarrow$ this results in solution of many Riemann problems

Near $x_{j+1/2}$:



For Δt sufficiently small, you can construct the solution exactly. Take

$$v_j^1 = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, \Delta t)\,dx$$

where $u(x, \Delta t)$ is the exact solution.

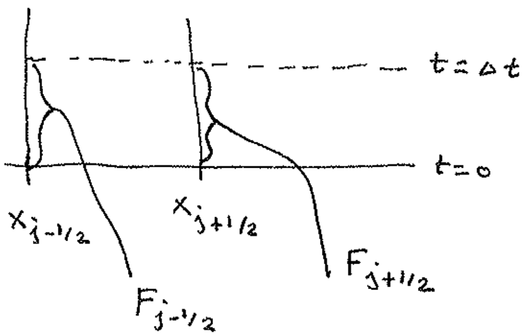
$$v_j^1 = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(x, \Delta t)\,dt = v_j^0 - \frac{\Delta t}{\Delta x}\left[F_{j+1/2}^0 - F_{j-1/2}^0\right]$$

From exact conservation law

$$v_j^1 = v_j^0 - \frac{\Delta t}{\Delta x}\left[F^0_{j+1/2} - F^0_{j-1/2} \right]$$

where $$F^0_{j+1/2} = \frac{1}{\Delta t}\int_0^{\Delta t} f(u(x_{j+1/2},t))\,dt$$

$$F^0_{j-1/2} = \frac{1}{\Delta t}\int_0^{\Delta t} f(u(x_{j-1/2},t))\,dt$$



for Δt sufficiently small u is constant along the interface between each Riemann problem.

$$\Rightarrow F_{j+1/2} = f(\underbrace{u(x_{j+1/2},t)}_{u^*(v_j^0,\, v_{j+1}^0)})$$

Godunov's Method

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left(f(u^*(v_j^n, v_{j+1}^n)) - f(u^*(v_{j-1}^n, v_j^n)) \right)$$

where $u^*(u_\ell, u_r)$ is the exact solution of the Riemann problem

$$u_t + f(u)_x = 0$$

$$u(x,0) = \begin{cases} u_\ell, & x<0 \\ u_r, & x>0 \end{cases}$$

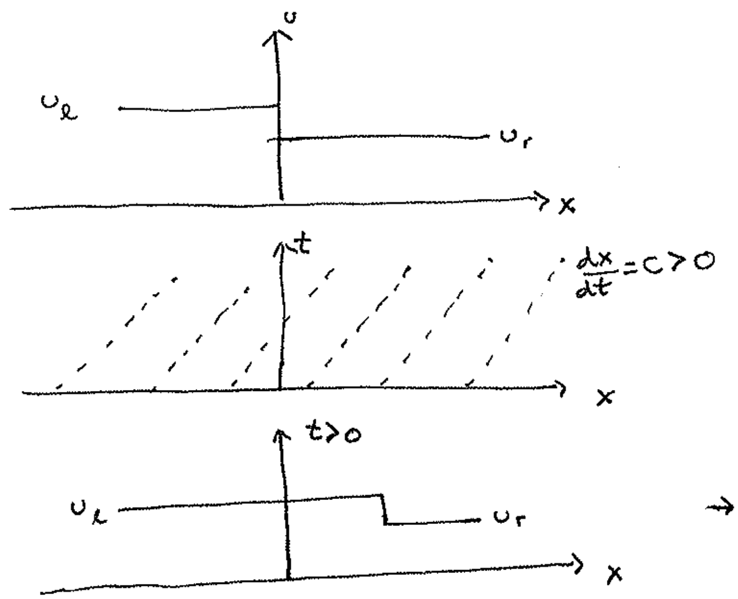
along $x=0$ for $t>0$

Apply the method for the linear advection equation:

$$u_t + c u_x = 0 \,, \quad c > 0 \,, \quad u(x,0) = u_0(x)$$

here, Flux, $f(u) = cu$

The Riemann problem is $u_t + c u_x = 0$, $u(x,0) = \begin{cases} u_\ell , & x<0 \\ u_r , & x>0 \end{cases}$



$\rightarrow \; u^*(u_\ell, u_r) = u_\ell$

$\rightarrow \; f(u^*(u_\ell, u_r)) = c u_\ell$

So Godunov's method becomes

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left(c v_j^n - c v_{j-1}^n \right)$$

$$= v_j^n - \frac{c\Delta t}{\Delta x}\left(v_j^n - v_{j-1}^n \right) \rightarrow \text{First-order upwind method}$$

For a general flux function, Godunov may be regarded as a nonlinear upwind method (first order accurate).

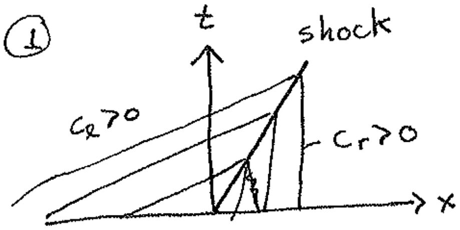
Now consider the Riemann problem for a nonlinear, scalar flux function, with $f''(u) \neq 0$, (a convex flux function). The characteristic speed is monotone if $f''(u) \neq 0$.

There are 4 cases to consider

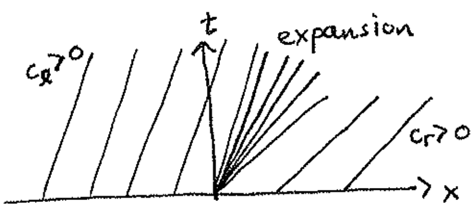
set $c_\ell = f'(u_\ell)$ - characteristic speed associated w/ left state

$c_r = f'(u_r)$ - characteristic speed associated w/ right state

---

① 

$\Rightarrow \quad c_\ell > c_r \geq 0$



$\Rightarrow \quad c_r > c_\ell \geq 0$

in both cases, $u^* = u_\ell$

---

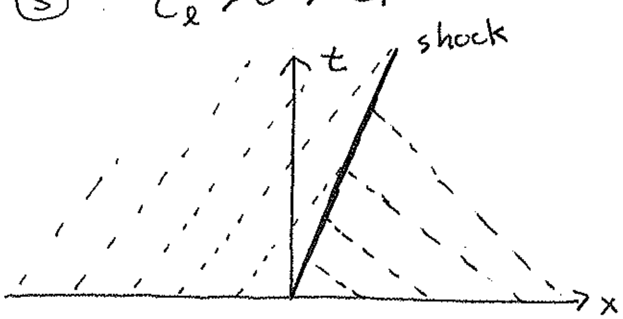
② $c_\ell < 0$, $c_r < 0$

these are the opposite pictures to case ① - shocks to the left, expansions to the left

$\Rightarrow \quad u^* = u_r$

---

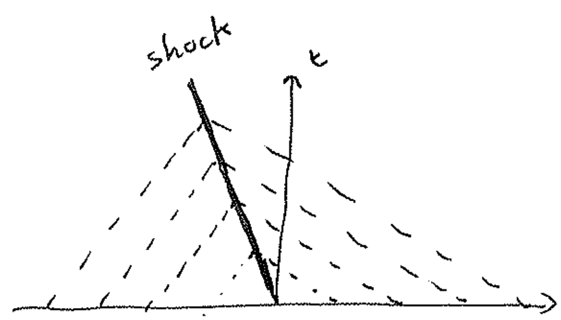
③ $c_\ell > 0 > c_r$



$u^* = u_\ell$

$$\frac{f(u_r) - f(u_\ell)}{u_r - u_\ell} > 0$$

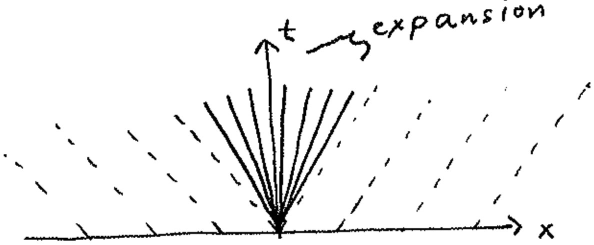
OR



$u^* = u_r$

$$\frac{f(u_r) - f(u_\ell)}{u_r - u_\ell} < 0$$

(4) $c_\ell < 0 < c_r$



hardest case because u^* is neither u_ℓ or u_r

Recall, the expansion solution is

$$u(x,t) = \begin{cases} u_\ell, & \frac{x}{t} \le c_\ell \\ z, & c_\ell < \frac{x}{t} < c_r \\ u_r, & \frac{x}{t} \ge c_r \end{cases}$$

where z solves $f'(z) = \frac{x}{t}$

$\rightarrow$ $\boxed{u^* \text{ solves } f'(u^*) = 0, \text{ "sonic case"}}$

You can show that all cases are covered by

$$f(u^*(u_\ell, u_r)) = \begin{cases} \min\limits_{u_\ell \le u \le u_r} f(u), & u_\ell \le u_r \quad \leftarrow \text{ inequalities} \\ \max\limits_{u_r \le u \le u_\ell} f(u), & u_r < u_\ell \quad \leftarrow \text{ strict inequalities} \end{cases}$$

Review of Big Picture: Scalar Conservation Laws

$u_t + f(u)_x = 0$, $|x| < \infty$, $t = 0$, $u(x,0) = u_0(x)$

- conservative Finite volume scheme

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left(F_{j+1/2}^n - F_{j-1/2}^n\right)$$

- centered methods
  - Lax Friedrichs, 1st order
  - Lax Wendroff, 2nd order
- upwind methods
  - Godunov, 1st order

1^{st}-order methods $\to$ dissipation near jumps

2^{nd}-order methods $\to$ dispersion near jumps

we want to discuss **high resolution methods**:

these are methods that are at least second order accurate but avoid oscillations near jumps, and where shock speeds are computed accurately.

in smooth regions of solution

High Resolution Methods

There are various approaches to obtain high resolution methods:

1) Flux limiters
2) slope limiters.

Flux Limiters (see Leveque 16.2 - thin green book)

begin with a conservative scheme

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left(G(v_j^n, v_{j+1}^n) - G(v_{j-1}^n, v_j^n)\right)$$

where $G(v_\ell, v_r)$ is a numerical flux function.

To develop a high resolution method, we would like

$G \simeq G_{HIGH}$ when v is smooth

and $G \simeq G_{LOW}$ near shocks

Set $G(u_\ell, u_r) = G_L(u_\ell, u_r) + \phi\left[G_H(u_\ell, u_r) - G_L(u_\ell, u_r)\right]$

where G_L is a low order flux (eg LF)

G_H is a high order flux (eg LW)

ϕ is a limiter.

want $\phi \approx 1$ when u is smooth

≈ 0 near shock

Consider the simple case of linear advection

$$u_t + c u_x = 0, \quad c > 0$$

L.W. : $$v_j^{n+1} = v_j^n - \frac{\sigma}{2}\left(v_{j+1}^n - v_{j-1}^n\right) + \frac{\sigma^2}{2}\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n\right), \quad \sigma = \frac{c\Delta t}{\Delta x}$$

this is usual way of writing LW, but rewrite as the following:

$$v_j^{n+1} = \underbrace{v_j^n - \sigma\left(v_j^n - v_{j-1}^n\right)}_{\substack{1^{st}\text{ order}\\ \text{upwind}}} - \underbrace{\frac{1}{2}\sigma(1-\sigma)\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n\right)}_{\substack{\text{flux correction to make}\\ \text{scheme } 2^{nd}\text{ order}}}$$

1st order upwind → stable if $0 \le \sigma \le 1$ →

For this range of σ, $\sigma(1-\sigma) > 0$ therefore this flux correction term is "anti-diffusive", to remove the diffusive error in upwind method

If the solution is smooth, then the anti-diffusive correction is effective, ie it makes the scheme 2^{nd} order. However, if the solution is not smooth, the correction overcorrects to give oscillations.

$$v_j^{n+1} = v_j^n - \sigma\left(v_j^n - v_{j-1}^n\right) - \frac{1}{2}\sigma(1-\sigma)\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n\right)$$

the corresponding flux in the conservative scheme

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left(G(v_j^n, v_{j+1}^n) - G(v_{j-1}^n, v_j^n)\right)$$

is
$$G(v_j^n, v_{j+1}^n) = \underbrace{+c v_j^n}_{G_L} + \underbrace{\frac{1}{2}c(1-\sigma)\left(v_{j+1}^n - v_j^n\right)}_{G_H - G_L}$$

Limited Flux :

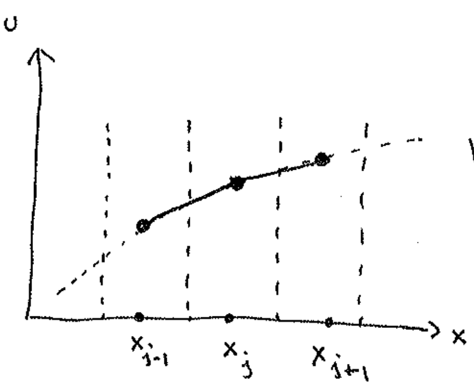
$$G(v_j^n, v_{j+1}^n) = c v_j^n + \frac{1}{2}c(1-\sigma)\left(v_{j+1}^n - v_j^n\right)\phi_j$$

limited

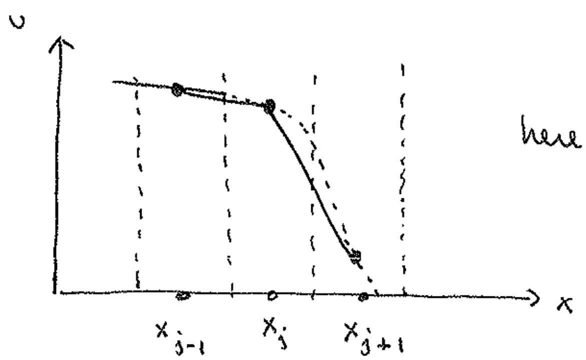
The limiter needs some measure of smoothness of solution

One choice :

$$\theta_j = \frac{v_j^n - v_{j-1}^n}{v_{j+1}^n - v_j^n} = \text{ratio of successive differences}$$



here, $\theta_j \approx 1$ $\rightarrow$ in smooth, monotone case, $\theta_j \approx 1$ (points of extrema do present problems)



here, $\theta_j \not\approx 1$

Set $\phi_j = \Phi(\theta_j)$

ie let ϕ_j be some function of θ_j

there are many such functions (Bee, Van Leer, etc...)

Need some way to guide the choice of limiters:

=> **TVD Methods** (total variation diminishing)

define the total variation of a grid function:

$$TV(v_j^n) \equiv \sum_{j=-\infty}^{\infty} |v_j^n - v_{j-1}^n|$$ → add up all successive differences in absolute value

usually assume that v_j approaches a constant as $j \to \pm\infty$ so that $TV(v_j)$ is bounded.

$$\boxed{\text{A TVD method is one in which } \quad TV(v_j^{n+1}) \leq TV(v_j^n) \quad \text{for all } n}$$

It turns out that solutions of the scalar conservation law have the continuous version of ~~this~~ TVD.

$$TV(u) = \int_{-\infty}^{\infty} |u_x| \, dx$$

so want to reproduce this behavior in numerical scheme.

For our purposes,

TVD will imply no oscillations.

We will construct limiter functions such that TVD occurs.

11/10/03

Flux Limiters (continued)

Consider $u_t + c u_x = 0$

Lax-Wendroff:

$$v_j^{n+1} = \underbrace{v_j^n - \sigma(v_j^n - v_{j-1}^n)}_{\substack{\text{upwind method} \\ 1^{st}\text{ order}}} - \underbrace{\frac{\sigma}{2}(1-\sigma)(v_{j+1}^n - 2v_j^n + v_{j-1}^n)}_{\substack{2^{nd}\text{ order} \\ \text{correction}}} \quad , \sigma = \frac{a\Delta t}{\Delta x}, 0<\sigma\leq 1$$

the flux with the limiter ϕ is

$$G(v_j^n, v_{j+1}^n) = c v_j^n + \frac{1}{2}c(1-\sigma)(v_{j+1}^n - v_j^n)\,\phi_j$$

Set the limiter to be some function of a grid function θ_j^n, where θ_j^n measures the smoothness of the solution

$$\phi_j^n = \Phi(\theta_j^n) \quad , \quad \theta_j^n = \frac{v_j^n - v_{j-1}^n}{v_{j+1}^n - v_j^n}$$

want to construct Φ st the resulting method is TVD.

Lax-Wendroff scheme with limiter

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left[c(v_j^n - v_{j-1}^n) + \frac{c}{2}(1-\sigma)\left((v_{j+1}^n - v_j^n)\phi_j^n - (v_j^n - v_{j-1}^n)\phi_{j-1}^n\right)\right]$$

May rewrite as

$$v_j^{n+1} = v_j^n - \left[\sigma - \frac{1}{2}\sigma(1-\sigma)\phi_{j-1}^n\right](v_j^n - v_{j-1}^n) - \left[\frac{\sigma}{2}(1-\sigma)\phi_j^n\right](v_{j+1}^n - v_j^n)$$

Theorem (by Harten)

In order for a method of the form

$$v_j^{n+1} = v_j^n - \alpha_{j-1}(v_j^n - v_{j-1}^n) + \beta_j(v_{j+1}^n - v_j^n)$$

to be TVD, the following conditions are sufficient

$$\alpha_{j-1} \geq 0 \quad , \quad \beta_j \geq 0 \quad , \quad \alpha_{j-1} + \beta_j \leq 1 \qquad \text{for all } j$$

For a proof, see Leveque, pg 178.

In our case, we have

$$\left.\begin{aligned} \alpha_{j-1} &= \sigma - \frac{\sigma}{2}(1-\sigma)\phi_{j-1}^n \\ \beta_j &= -\frac{\sigma}{2}(1-\sigma)\phi_j^n \end{aligned}\right\} \text{this selection is not unique.}$$

Notice, if $\phi_j^n \approx 1$, $\beta_j < 0$ because $0 \leq \sigma \leq 1$.

Another possibility is

$$\alpha_{j-1} = \sigma - \frac{1}{2}\sigma(1-\sigma)\phi_{j-1}^n + \frac{1}{2}\sigma(1-\sigma)\phi_j^n\left(\frac{v_{j+1}^n - v_j^n}{v_j^n - v_{j-1}^n}\right)$$

$$\beta_j = 0$$

Then the sufficient conditions for TVD become

$$0 \leq \alpha_{j-1} \leq 1$$

$$\Rightarrow \quad 0 \leq \sigma\left[1 + \frac{1}{2}(1-\sigma)\left(\underbrace{\phi_j^n}_{\Phi(\theta_j^n)}\underbrace{\left(\frac{v_{j+1}^n - v_j^n}{v_j^n - v_{j-1}^n}\right)}_{\frac{1}{\theta_j^n}} - \underbrace{\phi_{j-1}^n}_{\Phi(\theta_{j-1}^n)}\right)\right] \leq 1$$

$$\Rightarrow \quad 0 \leq 1 + \frac{1}{2}(1-\sigma)\left(\frac{\Phi(\theta_j^n)}{\theta_j^n} - \Phi(\theta_{j-1}^n)\right) \leq 1$$

You can show under what conditions this inequality is satisfied

$$0 \le 1 + \frac{1}{2}(1-\sigma)\left(\frac{\Phi(\theta_j^n)}{\theta_j^n} - \Phi(\theta_{j-1}^n) \right) \le 1$$

This condition is satisfied if $0 \le \sigma \le 1$

and if

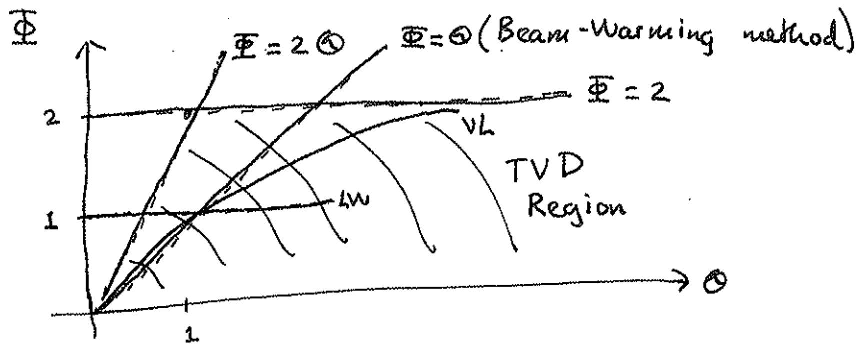
$$\left| \frac{\Phi(\theta_j^n)}{\theta_j^n} - \Phi(\theta_{j-1}^n) \right| \le 2 \qquad , \quad \theta_j^n = \frac{v_j^n - v_{j-1}^n}{v_{j+1}^n - v_j^n}$$

Notice, if $\theta_j^n < 0$, then neighboring differences in v_j^n have opposite sign . . . , ...

Safest to set $\Phi(\theta_j^n) = 0$ when $\theta_j^n < 0$. This avoids the growth of peaks or dips, but also makes the scheme only first order accurate near smooth local extrema.

So to satisfy the above condition, let

$$0 \le \frac{\Phi(\theta_j^n)}{\theta_j^n} \le 2 \quad , \qquad 0 \le \Phi(\theta_i^n) \le 2$$



Note, $\Phi = 1$ for Lax-Wendroff
$\Phi = \theta$ for Beam-Warming method

A requirement for 2^{nd} order is that $\Phi(1) = 1$, smoothly.

We can write down many functions that stay in TVD region and pass through $\Phi(1) = 1$, smoothly.

van Leer limiter $\quad \Phi = \dfrac{|\theta| + \theta}{1 + |\theta|} \Rightarrow$ TVD

Example

$u_t + u_x = 0 \quad , \quad t > 0 \quad (c = 1)$

$u(x,0) = \tanh(\lambda x) \quad , \quad \lambda$ - parameter

exact solution is $u(x,t) = \tanh(\lambda(x-t))$

Choose a grid $x_j = j\Delta x$ and choose large enough range of j st $u \to$ constant for $|j|$ large.

Integrate numerically using the Flux-limited Lax-Wendroff method

$$\phi_j = \Phi(\theta_j^n) \quad , \quad \Phi(\theta_j^n) = \frac{|\theta| + \theta}{1 + |\theta|}$$

$$\theta_j^n = \frac{v_j^n - v_{j-1}^n}{v_{j+1}^n - v_j^n} = \frac{\delta_{-x} v_j^n}{\delta_{+x} v_j^n}$$

sgn(x)

$$\Rightarrow \quad \phi_j^n = \frac{\left|\frac{\delta_{-x} v_j^n}{\delta_{+x} v_j^n}\right| + \frac{\delta_- v_j^n}{\delta_+ v_j^n}}{1 + \left|\frac{\delta_{-x} v_j^n}{\delta_{+x} v_j^n}\right|} = \frac{|\delta_{-x} v_j^n| + \delta_- v_j^n \frac{\delta_+ v_j^n}{|\delta_+ v_j^n|}}{|\delta_+ v_j^n| + |\delta_- v_j^n|}$$

$$\rightarrow \quad \phi_j^n = \frac{|\delta_{-x} v_j^n| + \delta_- v_j^n \cdot \operatorname{sgn}(\delta_{+x} v_j^n)}{|\delta_{+x} v_j^n| + |\delta_- v_j^n| + \epsilon} \quad , \quad \epsilon \sim 10^{-10}$$

this avoids singularity as you approach $u \sim$ const for $|x|$ large

Slope Limiters

- this approach is based on Godunov's method

$$u_t + f(u)_x = 0$$

Godunov's method (1st order upwind method)

$$v_j^{n+1} = v_j^n - \frac{\Delta t}{\Delta x}\left(F_{j+1/2}^n - F_{j-1/2}^n \right)$$

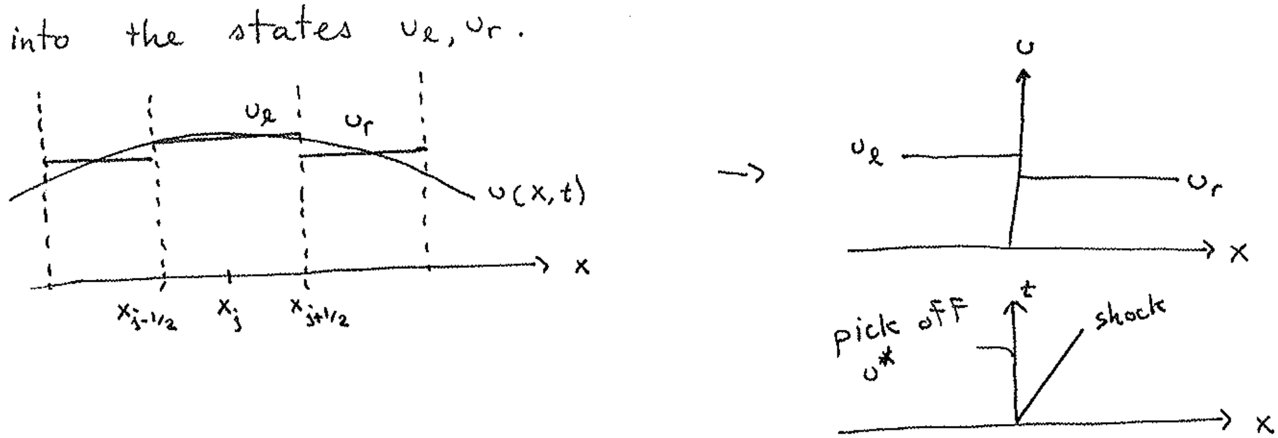
where $F_{j+1/2}^n = f(u^*(v_j^n, v_{j+1}^n))$

the state $u^*(u_\ell, u_r)$ is found by solving a Riemann problem

$$u_t + f(u)_x = 0 \, , \quad u(x,0) = \begin{cases} u_\ell & , x<0 \\ u_r & , x>0 \end{cases}$$

then $u^* = u(x,t)$ for $x=0$, $t>0$.

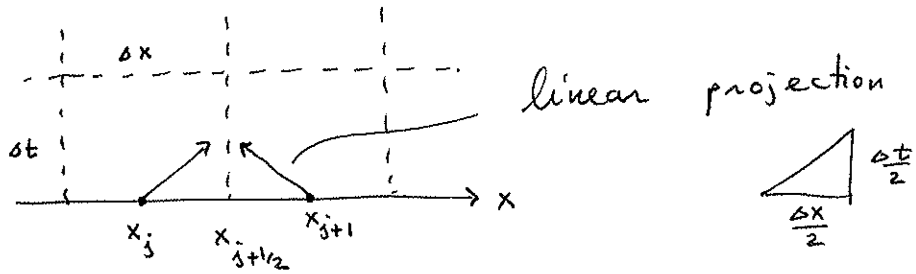
To obtain high resolution method, must somehow increase the order of accuracy. In order to do this, you need to include some derivative information into the states u_ℓ, u_r.



To increase accuracy, instead of using piecewise constant approximations, use piecewise linear approximations, (or quadratic, cubic, etc...).

slope limiters

For second-order approximations, need to include slope information → think midpoint rule.



use Taylor series

$$u\left(x_j + \frac{\Delta x}{2}, t_n + \frac{\Delta t}{2}\right) = u(x_j, t_n) + \frac{\Delta x}{2} u_x(x_j, t_n) + \frac{\Delta t}{2} u_t(x_j, t_n) + \cdots$$

$$= u + \frac{\Delta x}{2} u_x - \frac{\Delta t}{2} f'(u) u_x + \cdots$$

$$= u + \underbrace{\frac{1}{2}\left(1 - \frac{\Delta t}{\Delta x} f'(u)\right) \Delta x\, u_x}_{\text{slope correction}} + \cdots$$

For the Riemann problem about $x_{j+1/2}$, take

$$u_\ell = v_j^n + \frac{1}{2}\left(1 - \frac{\Delta t}{\Delta x} f'(v_j^n)\right) \delta_{+x} v_j^n$$

$$u_r = v_{j+1}^n - \frac{1}{2}\left(1 + \frac{\Delta t}{\Delta x} f'(v_{j+1}^n)\right) \delta_{-x} v_j^n$$

this results in a $2^{\underline{nd}}$ order extension of Godunov's method

There is no limiting in these formulas so that oscillations near shocks would occurr. To supress oscillation we want to include a slope limiter.

One effective choice:

- $u_\ell = v_j^n + \frac{1}{2}\left(1 - \frac{\Delta t}{\Delta x}\max\left(f'(v_j^n), 0\right)\right)\Delta v_j^n$

$$\Delta v_j^n = \text{minmod}\left(\delta_{+x} v_j^n \ , \ \delta_{-x} v_j^n\right)$$

where the minimum modulus function is defined by

$$\text{minmod}(a, b) = \begin{cases} a & \text{if } ab > 0 \ , \ |a| < |b| \\ b & \text{if } ab > 0 \ , \ |a| > |b| \\ 0 & \text{otherwise} \end{cases}$$

- $u_r = v_{j+1}^n - \frac{1}{2}\left(1 + \frac{\Delta t}{\Delta x}\min\left(f'(v_{j+1}^n), 0\right)\right)\Delta^n v_{j+1}$

11/13/03

Systems of Conservation Laws

$$\underline{u}_t + \underline{f}(\underline{u})_x = 0$$

where $\underline{u}$ is a vector of m state variables and $\underline{f}$ is a vector of m flux functions. The system is hyperbolic if $\underline{f}_{\underline{u}}(\underline{u})$ is diagonalizable with real eigenvalues, $\lambda_1(\underline{u}) \le \lambda_2(\underline{u}) \le \dots \le \lambda_m(\underline{u})$

Finite volume method

$$\underline{v}_j^{n+1} = \underline{v}_j^n - \frac{\Delta t}{\Delta x}\left(\underline{F}_{j+1/2}^n - \underline{F}_{j-1/2}^n\right)$$

Here
$$\underline{v}_j^n \approx \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \underline{u}(x, t_n)\, dx$$

where $\underline{F}_{j+1/2}^n$ is a numerical flux function

Standard methods such as Lax-Friedrichs and Lax-Wendroff carry over

$$\underline{F}^n_{j+1/2} = \frac{1}{2}\left(\underline{f}(\underline{v}^n_j) + \underline{f}(\underline{v}^n_{j+1})\right) - \frac{\Delta x}{2\Delta t}\left(\underline{v}^n_{j+1} - \underline{v}^n_j\right) \quad \text{- Lax-Friedrichs}$$

$$\underline{F}^n_{j+1/2} = \frac{1}{2}\left(\underline{f}(\underline{v}^n_j) + \underline{f}(\underline{v}^n_{j+1})\right) - \frac{\Delta t}{2\Delta x}\,\underline{f}_{\underline{u}}(\underline{v})\left(\underline{f}(\underline{v}^n_{j+1}) - \underline{f}(\underline{v}^n_j)\right) \quad \text{- Lax-Wendroff}$$

$$\uparrow$$

$$\frac{1}{2}(\underline{v}^n_j + \underline{v}^n_{j+1})$$

Godunov's method

$$\underline{F}^n_{j+1/2} = \underline{f}\left(\underline{u}^*(\underline{v}^n_j, \underline{v}^n_{j+1})\right)$$

where $\underline{u}^*(\underline{u}_\ell, \underline{u}_r)$ is found by solving the Riemann problem

$$\underline{u}_t + \underline{f}(\underline{u})_x = 0$$

$$\underline{u}(x,0) = \begin{cases} \underline{u}_\ell, & x<0 \\ \underline{u}_r, & x>0 \end{cases}$$

take $\underline{u}^* = \underline{u}(0,t)$, $t>0$

The solution of the Riemann problem for systems is more complicated and more costly computationally. Often solve an approximate Riemann problem instead. Usually these are based on some linearization,

$$\underline{u}_t + \underline{\underline{A}}\,\underline{u}_x = 0, \quad \underline{u}(x,0) = \begin{cases} \underline{u}_\ell, & x<0 \\ \underline{u}_r, & x>0 \end{cases}$$

where $\underline{\underline{A}}$ approximates the Jacobian $\underline{f}_{\underline{u}}$. You require

$$\underline{\underline{A}} \to \underline{f}_{\underline{u}}(\underline{u}) \quad \text{as} \quad \underline{u}_\ell \to \underline{u} \quad \text{and} \quad \underline{u}_r \to \underline{u}$$

Typically, take $\underline{\underline{A}}$ to be the Jacobian evaluated at some average state $\hat{u}$, $\quad \underline{\underline{A}} = \underline{f}_{\underline{u}}(\hat{\underline{u}})$, where $\hat{\underline{u}} = \frac{\underline{u}_l + \underline{u}_r}{2}$

this is Roe's method.

The solution of the linear problem is straightforward.

Compute the eigenvalues of $\underline{\underline{A}}$, (because these are the characteristic speeds of the linear problem) and the corresponding eigenvectors:

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_m$$

$$\underline{r}_1, \underline{r}_2, \ldots, \underline{r}_m$$

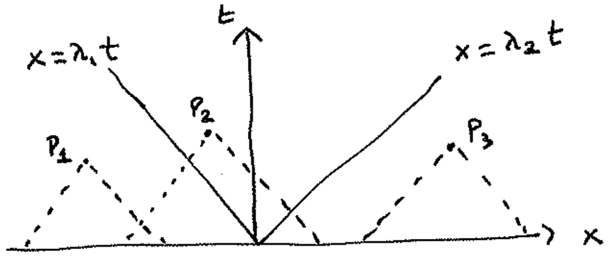
compute $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that

$$\alpha_1 \underline{r}_1 + \alpha_2 \underline{r}_2 + \cdots + \alpha_m \underline{r}_m = \underline{u}_r - \underline{u}_l$$

Then the solution becomes

$$\boxed{\begin{aligned} \underline{u}(x,t) &= \underline{u}_l + \sum_{\lambda_k < \frac{x}{t}} \alpha_k \underline{r}_k \\ &= \underline{u}_r - \sum_{\alpha_k > \frac{x}{t}} \alpha_k \underline{r}_k \end{aligned}}$$

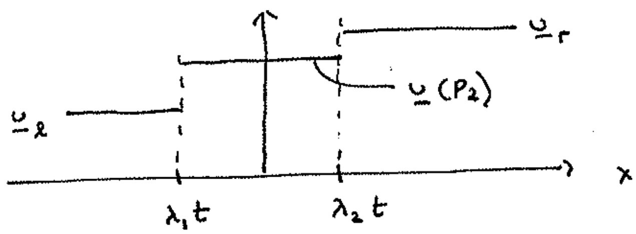
Example, $m=2$, $\lambda_1 < 0$, $\lambda_2 > 0$



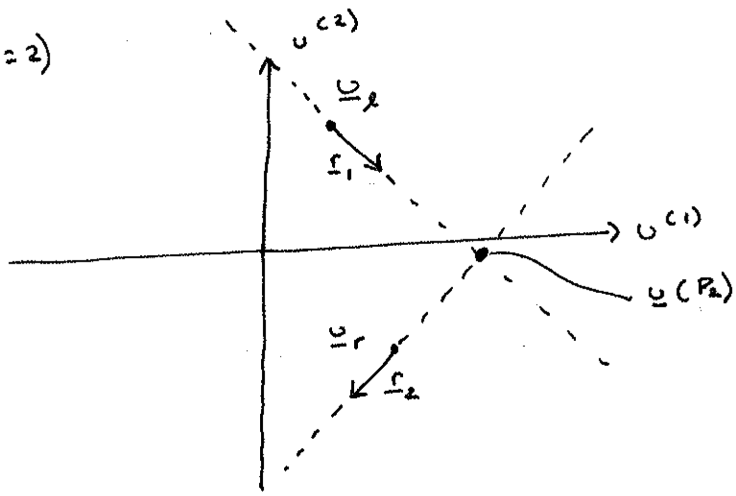
$\underline{u}(P_1) = \underline{u}_\ell$

$\underline{u}(P_2) = \underline{u}_\ell + \alpha_1 \underline{r}_1$

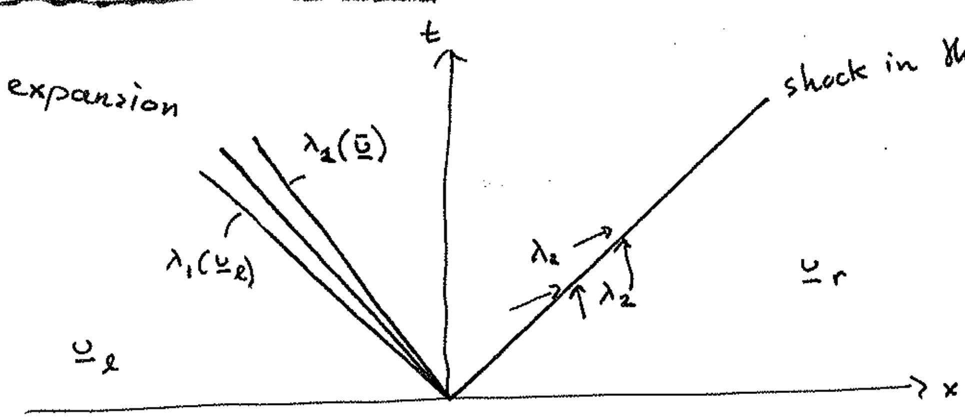
$\underline{u}(P_2) = \underline{u}_\ell + \alpha_1 \underline{r}_1 + \alpha_2 \underline{r}_2 = \underline{u}_r$

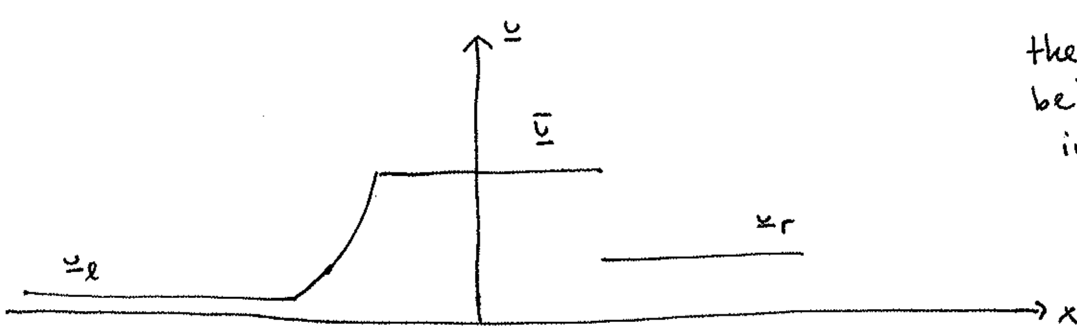


Phase plane: $(m=2)$



For the nonlinear case, an example is the following, for $m=2$





you don't know the structure beforehand, typical involves iteration

Multiple Dimensions

Consider two-dimensions

$$u_t + f(u)_x + g(u)_y = 0 \quad , \quad u(x,y,0) = u_0(x,y)$$

where $u(x,y,t)$ is the state variable and $f(u), g(u)$ are fluxes in the x, y directions, respectively.

Finite volume scheme - obtained by integrating this equation in x, y, t. Grid: $x_j = j\Delta x$, $y = k \Delta y$

$$v_{j,k}^{n} = \frac{1}{\Delta x \Delta y} \int_{x_{j-1/2}}^{x_{j+1/2}} \int_{y_{k-1/2}}^{y_{k+1/2}} u(x,y,t_n)\, dy\, dx$$

Integrate on a cell to get

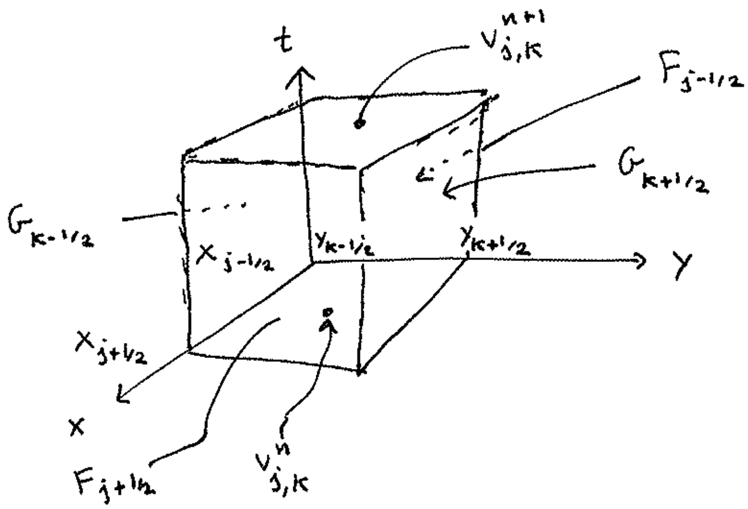
$$\boxed{v_{j,k}^{n+1} = v_{j,k}^{n} - \frac{\Delta t}{\Delta x}\left(F_{j+1/2,k}^{n} - F_{j-1/2,k}^{n} \right) - \frac{\Delta t}{\Delta y}\left(G_{j,k+1/2}^{n} - G_{j,k-1/2}^{n} \right)}$$

where

$$F_{j+1/2,k}^{n} = \frac{1}{\Delta y \Delta t} \int_{t_n}^{t_{n+1}} \int_{y_{k-1/2}}^{y_{k+1/2}} f\left(u(x_{j+1/2}, y, t)\right) dy\, dt$$

$$G_{j,k+1/2}^{n} = \frac{1}{\Delta x \Delta t} \int_{t_n}^{t_{n+1}} \int_{x_{j-1/2}}^{x_{j+1/2}} g\left(u(x, y_{k+1/2}, t)\right) dx\, dt$$

referred to as an unsplit scheme - there are also directional splitting



Lax-Friedrichs formulas

$$F^n_{j+1/2,k} = \frac{1}{2}\left(f(v^n_{j,k}) + f(v^n_{j+1,k}) \right) - \frac{\Delta x}{2\Delta t}\left(v^n_{j+1,k} - v^n_{j,k} \right)$$

$$G^n_{j,k+1/2} = \frac{1}{2}\left(g(v^n_{j,k}) + g(v^n_{j,k+1}) \right) - \frac{\Delta y}{2\Delta t}\left(v^n_{j,k+1} - v^n_{j,k} \right)$$

Directional Splitting

General scheme, $v^{n+1}_{j,k} = S_y(\Delta t)\, S_x(\Delta t)\, v^n_{j,k}$

where $S_x(\Delta t)$ denotes one time step of a scheme to solve $u_t + f(u)_x = 0$, a 1D scheme. Then $S_y(\Delta t)$ denotes one time step of a scheme to solve $u_t + g(u)_y = 0$. This is only first order accurate in time, no matter how you solve each individual 1D problem. This can be avoided by implementing Strang splitting.

Strang splitting: $v^{n+1}_{j,k} = S_x\left(\frac{\Delta t}{2}\right) S_y(\Delta t)\, S_x\left(\frac{\Delta t}{2}\right) v^n_{j,k}$

$\rightarrow$ 2nd order accurate (at most)

Strang splitting: $v_{j,k}^{n+1} = S_x\left(\frac{\Delta t}{2}\right) S_y(\Delta t)\, S_x\left(\frac{\Delta t}{2}\right)$

For many steps:

$$v_{j,k}^{n+2} = \left[S_x\left(\tfrac{\Delta t}{2}\right) S_y(\Delta t)\, S_x\left(\tfrac{\Delta t}{2}\right)\right]\left[S_x\left(\tfrac{\Delta t}{2}\right) S_y(\Delta t)\, S_x\left(\tfrac{\Delta t}{2}\right)\right]$$

$$\underbrace{S_x\left(\tfrac{\Delta t}{2}\right) S_x\left(\tfrac{\Delta t}{2}\right)}_{S_x(\Delta t)}$$

becomes more efficient

CFL condition becomes something like:

$$\frac{\Delta t}{\Delta x} \max(\lambda_p) + \frac{\Delta t}{\Delta y} \max(\nu_p) \le CFL$$

where λ_p eigenvalues of f_u
ν_p eigenvalues of g_u

ELLIPTIC EQUATIONS

Recall the 2nd order PDE

$$A u_{xx} + 2B u_{xy} + C u_{yy} = D$$

The equation is elliptic if $B^2 - AC < 0$

Coordinate transform leads to $u_{\xi\xi} + u_{\eta\eta} = \hat{D}$, Poisson's eqn

or consider, $u_{xx} + u_{yy} = f(x, y, u, u_x, u_y)$

More generally, in k dimensions, the PDE

$$\sum_{p,q=1}^{k} a_{p,q}(\underline{x}) \frac{\partial^2 u}{\partial x_p \partial x_q} + \sum_{p=1}^{k} b_p(\underline{x}) \frac{\partial u}{\partial x_p} + c(\underline{x})\, u = d(\underline{x})$$

with $a_{p,q} = a_{q,p}$, is elliptic if $\underline{\underline{A}} = [a_{p,q}(\underline{x})]$ is positive definite.

Elliptic Eqns

Canonical forms

$-u_{xx} - u_{yy} = F(x,y)$ Poisson's Eq

$-u_{xx} - u_{yy} = 0$ Laplace eqn

boundary value problem

Some properties of Laplace's equation

$$u_{xx} + u_{yy} = 0 \quad , \quad x^2+y^2 < 1$$

$$u = \text{given on } x^2+y^2 = 1 \qquad \text{(Dirichlet)}$$

Formulate in polar coordinates

$$\frac{1}{r}(r u_r)_r + \frac{1}{r^2} u_{\theta\theta} = 0 \quad , \quad 0<r<1 \quad , \quad 0 \le \theta \le 2\pi$$

$$u(1,\theta) = f(\theta)$$

Solution via separation of variables

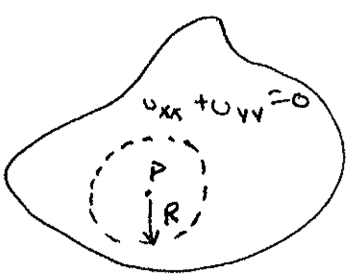
$$u(r,\theta) = a_0 + \sum_{n=1}^{\infty} r^n \left[a_n \cos n\theta + b_n \sin n\theta\right]$$

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\, d\theta$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} f(\theta) \cos n\theta \, d\theta$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} f(\theta) \sin n\theta \, d\theta$$

Notice that at $r=0$, $u(r,\theta) = a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)\,d\theta$, which is the mean value of the solution on the perimeter. The same property holds more generally.



Mean-value property for Laplace's Eqn

u_p = mean value of u on circle $r=R$

this holds for any point P and any radius R provided you stay in domain

Useful to prove maximum principle, uniqueness, ...

Max / Min Principle

If $u(x,y)$ satisfies Laplace's equation for a domain Ω, then the max (or min) of u must occur on $\partial\Omega$. Proof by contradiction and uses the mean value property.

Uniqueness of Solution (for Dirichlet problem)

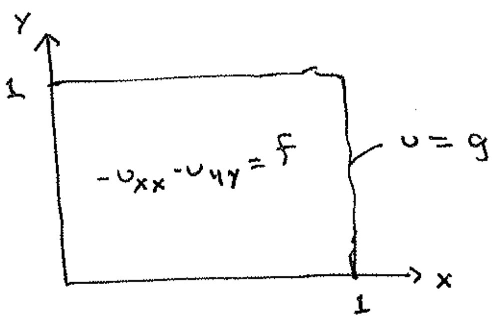
$$u_{xx} + u_{yy} = 0 \qquad , \; x, y \in \Omega$$
$$u = \text{given} \qquad x, y, \in \partial\Omega$$

Consider the difference of two solutions ~~to~~ and use max/min principle to show that the difference is identically zero.

Numerics

Model problem: $\quad -u_{xx} - u_{yy} = f(x,y) \qquad 0 < x < 1, \; 0 < y < 1$

$\qquad u = g(x,y)$ on boundary



Solve using finite differences:

grid $\quad x_j = j\Delta x \; , \quad \Delta x = 1/N$

$\qquad\; y_k = k\Delta y \; , \quad \Delta y = 1/M$

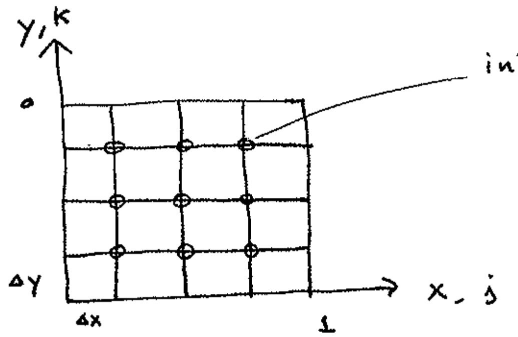
Set $\quad v_{j,k} \simeq u(x_j, y_k)$

Replace u_{xx} and u_{yy} by centered differences:

$$\left(-\frac{1}{\Delta x^2}\delta_x^2 - \frac{1}{\Delta y^2}\delta_y^2\right) v_{j,k} = F(x_j, y_k) \quad \text{for} \quad \begin{matrix} 1 \le j \le N-1 \\ 1 \le k \le M-1 \end{matrix}$$

Set $v_{j,k} = g(x_j, y_k)$ for (j,k) on the boundary

Example suppose $N = M = 4$



interior grid points

we have $(N-1)(M-1)$ linear equations for $v_{j,k}$ in the interior grid points

these linear systems result in a system

$$\underline{\underline{A}}\,\underline{v} = \underline{c}$$

we have to decide how to arrange $\underline{v}$. The question is at what grid point to start and then what direction do you solve in?

Let $\underline{v} = \begin{bmatrix} v_{11} & v_{21} & v_{31} & v_{12} & v_{22} & v_{32} & v_{13} & v_{23} & v_{33} \end{bmatrix}^T$

$$\underline{v} = \begin{bmatrix} v_{11} \\ v_{21} \\ v_{31} \\ v_{12} \\ v_{22} \\ v_{32} \\ v_{13} \\ v_{23} \\ v_{33} \end{bmatrix} \qquad \underline{\underline{A}} = \left[\begin{array}{ccc|ccc|ccc} x & x & & x & & & & & \\ x & x & x & & x & & & & \\ & x & x & & & x & & & \\ \hline x & & & x & x & & x & & \\ & x & & x & x & x & & x & \\ & & x & & x & x & & & x \\ \hline & & & x & & & x & x & \\ & & & & x & & x & x & x \\ & & & & & x & & x & x \end{array}\right] \qquad \underline{c} = \left[\begin{array}{c} x \\ x \\ x \\ \hline x \\ 0 \\ x \\ \hline x \\ x \\ x \end{array}\right]$$

where x - nonzero value

In general

1st grid line

$$\underline{v} = \left[\begin{array}{c} \left.\begin{array}{c} v_{11} \\ v_{21} \\ \vdots \\ v_{N-1,1} \end{array}\right\} N-1 \\ \hdashline \left.\begin{array}{c} v_{12} \\ \vdots \\ v_{N-1,2} \end{array}\right\} N-1 \\ \hdashline \vdots \\ \hdashline \left.\begin{array}{c} v_{1,M-1} \\ \vdots \\ v_{N-1,M-1} \end{array}\right\} N-1 \end{array}\right] \quad (M-1)\times(N-1)$$

2nd grid line

$\underline{\underline{A}}$ is $(N-1)(M-1) \times (N-1)(M-1)$

$$\underline{\underline{A}} = \begin{bmatrix} B & D & & & & \\ D & B & D & & & \\ & D & B & D & & \\ & & \ddots & \ddots & \ddots & \\ & & & D & B & D \\ & & & & D & B \end{bmatrix}$$

where $\underline{\underline{B}}$ =

tridiagonal $(N-1)\times(N-1)$

$$\underline{\underline{B}} = \begin{bmatrix} \frac{2}{\Delta x^2} + \frac{2}{\Delta y^2} & -\frac{1}{\Delta x^2} & & & \\ -\frac{1}{\Delta x^2} & \frac{2}{\Delta x^2} + \frac{2}{\Delta y^2} & -\frac{1}{\Delta x^2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{\Delta x^2} & \frac{2}{\Delta x^2} + \frac{2}{\Delta y^2} & -\frac{1}{\Delta x^2} \\ & & & -\frac{1}{\Delta x^2} & \frac{2}{\Delta x^2} + \frac{2}{\Delta y^2} \end{bmatrix}$$

$$\underline{\underline{D}} = \frac{-1}{\Delta y^2}\,\underline{\underline{I}}$$

$\underline{c}$ is a vector of boundary contributions

Solvability

Can we solve the linear system $\underline{\underline{A}}\,\underline{v} = \underline{c}$?
For this problem, you can show that $\underline{\underline{A}}$ is symmetric and positive definite, therefore $\underline{\underline{A}}$ is nonsingular and the system $\underline{\underline{A}}\,\underline{v} = \underline{c}$ has a unique solution for any choice of $\underline{c}$.

Consider the eigenvalue problem:

$$\underline{\underline{A}}\,\underline{w} = \lambda\,\underline{w}$$

and show that $\lambda > 0$.

$$\rightarrow \quad \left(-\frac{1}{\Delta x^2}\delta_x^2 - \frac{1}{\Delta y^2}\delta_y^2\right) w_{j,k} = \lambda w_{j,k} \qquad \text{for} \quad \begin{array}{l} 1 \le j \le N-1 \\ 1 \le k \le M-1 \end{array}$$

with $w_{j,k} = 0$ on boundary. The equation is a constant coefficient difference equations, which implies that you can apply separation of variables: $\quad w_{j,k} = a^j b^k \quad , \quad a, b = \text{constant}$

$$\begin{aligned} \rightarrow \quad \delta_x^2\, a^j b^k &= b^k \delta_x^2 a^j \\ &= b^k\left(a^{j-1} - 2a^j + a^{j+1}\right) \\ &= a^j b^k\left(\frac{1}{a} - 2 + a\right) \end{aligned}$$

$$\Rightarrow \quad -a^j b^k\left[\frac{1}{\Delta x^2}\left(\frac{1}{a} - 2 + a\right) + \frac{1}{\Delta y^2}\left(\frac{1}{b} - 2 + b\right)\right] = \lambda a^j b^k$$

$$\rightarrow \quad \boxed{\lambda = -\frac{1}{\Delta x^2}\left(\frac{1}{a} - 2 + a\right) - \frac{1}{\Delta y^2}\left(\frac{1}{b} - 2 + b\right)}$$

trick: assign $\quad \frac{1}{a} - 2 + a = -2 + 2\cos\theta$

and $\quad \frac{1}{b} - 2 + b = -2 + 2\cos\phi$

$\rightarrow$ ~~XXX~~ $\quad \frac{1}{a} + a = 2\cos\theta \quad \rightarrow \quad a^2 - 2a\cos\theta + 1 = 0$

$$a = \frac{\cancel{4}\cos\theta \pm \sqrt{4\cancel{a}\cos^2\theta - 4}}{\cancel{1}\;2}$$

$$\rightarrow \quad a = e^{\pm i\theta}$$

$$a = \cos\theta \pm \cancel{i}\sqrt{\cos^2\theta - 1}$$

$$\cos\theta \pm i\sin\theta$$

and likewise, $b = e^{\pm i\phi}$

$\Rightarrow$ solutions are a linear combination of

$$\left(e^{\pm i\theta}\right)^j, \quad \left(e^{\pm i\phi}\right)^k$$

use

$$w_{j,k} = c\left(e^{ij\theta} - e^{-ij\theta}\right)\left(e^{ik\phi} - e^{-ik\phi}\right), \quad c = \text{constant}$$

which satisfies BCs at $j = 0$ or $k = 0$ ($w_{j,k} = 0$ on $j = 0, k = 0$)

$\Rightarrow \quad w_{j,k} = c\,\sin(j\theta)\,\sin(k\phi)$ (constant c absorbed factors from $\sin = \frac{e^{()} - e^{-()}}{2}$)

note, $w_{j,k} = 0$ at $j = N \rightarrow \sin N\theta = 0 \rightarrow N\theta = p\pi$

$$\rightarrow \boxed{\theta = \frac{p\pi}{N}, \quad p = 1, \ldots, N-1}$$

and $w_{j,k} = 0$ at $k = M \rightarrow \boxed{\phi = \frac{q\pi}{M}, \quad q = 1, \ldots, M-1}$

Then we have

$$\lambda = \frac{-1}{\Delta x^2}\underbrace{\left(\frac{1}{a} - 2 + a\right)}_{-2 + 2\cos\theta} - \frac{1}{\Delta y^2}\underbrace{\left(\frac{1}{b} - 2 + b\right)}_{-2 + 2\cos\phi}$$

$$\lambda = \frac{2}{\Delta x^2}\underbrace{\left(1 - \cos\theta\right)}_{>0} + \frac{2}{\Delta y^2}\underbrace{\left(1 - \cos\phi\right)}_{>0}$$

$\Rightarrow$ $\boxed{\lambda > 0, \text{ strictly greater than zero; therefore } \underline{\underline{A}} \text{ is positive definite}}$

Solvability For More General Cases

Often can show solvability for linear systems obtained via finite differences for elliptic PDE by noting diagonal dominance.

Def: A matrix $\underline{\underline{A}}$ is diagonally dominant if

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j\neq i}}^{N} |a_{ij}| \quad \text{for all } i$$

and strictly diagonally dominant if

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{N} |a_{ij}| \quad \text{for all } i$$

$\underline{\underline{A}}$ is an $N \times N$ matrix

Thm: If $\underline{\underline{A}}$ is strictly diagonally dominant, then it is nonsingular.

consider $\underline{\underline{A}}\,\underline{x} = \underline{0} \Rightarrow \underline{x} = \underline{0}$ is the only solution

Example: $-u_{xx} - u_{yy} + \underbrace{a u_x + b u_y + c u}_{\text{lower order terms}} = d(x,y)$

Approximate using

$$\left(-\frac{1}{\Delta x^2}\delta_x^2 - \frac{1}{\Delta y^2}\delta_y^2 + \frac{a}{2\Delta x}\delta_{0x} + \frac{b}{2\Delta y}\delta_{0y} + c\right)v_{j,k} = d(x_j, y_k)$$

domain: $0<x<1$, $0<y<1$ with $u=g$ on the boundary

Question: Is this problem solvable?

The matrix $\underline{\underline{A}}$ is not necessarily positive definite or symmetric, so we examine the diagonal and off-diagonal elements.

diagonal element: $\frac{2}{\Delta x^2} + \frac{2}{\Delta y^2} + c > 0$ for

$\Delta x^2, \Delta y^2$ sufficiently small

sum of the magnitude off-diagonal elements

$$S = \left| -\frac{1}{\Delta x^2} + \frac{a}{2\Delta x} \right| + \left| \frac{-1}{\Delta x^2} - \frac{a}{2\Delta x} \right| + \left| \frac{-1}{\Delta y^2} + \frac{b}{2\Delta y} \right| + \left| \frac{-1}{\Delta y^2} - \frac{b}{2\Delta y} \right|$$

Suppose Δx and Δy are small enough so that

$$\frac{1}{\Delta x^2} \geq \frac{|a|}{2\Delta x} \Rightarrow \Delta x \leq \frac{2}{|a|} \Rightarrow \boxed{|a|\,\Delta x \leq 2}$$

$$\frac{1}{\Delta y^2} \geq \frac{|b|}{2\Delta y} \Rightarrow \Delta y \leq \frac{2}{|b|} \Rightarrow \boxed{|b|\,\Delta y \leq 2}$$

Notice a and b represent convective terms so therefore this is kind of like a CFL condition.

IF $|a|\Delta x \leq 2$ and $|b|\Delta y \leq 2$ then

$$S = \left(\frac{1}{\Delta x^2} - \cancel{\frac{a}{2\Delta x}}\right) + \left(\frac{1}{\Delta x^2} + \cancel{\frac{a}{2\Delta x}}\right) + \left(\frac{1}{\Delta y^2} - \cancel{\frac{b}{2\Delta y}}\right) + \left(\frac{1}{\Delta y^2} + \cancel{\frac{b}{2\Delta y}}\right)$$

$$\rightarrow \quad S = \frac{2}{\Delta x^2} + \frac{2}{\Delta y^2}$$

IF $c > 0$ then the linear system is strictly diagonally dominant $\Rightarrow$ solvable.

Suppose $c = 0$ then the problem is $\underbrace{\text{diagonally dominant}}_{\text{this is ok, just requires a little more work}}$.

IF $c < 0$, you may be singular.

If $c = 0$, then the system is only diagonally dominant, assuming $|a|\Delta x \le 2$, $|b|\Delta y \le 2$.

Def: An $N \times N$ matrix $\underline{\underline{A}}$ is reducible if either

a) $N = 1$ and $\underline{\underline{A}} = 0$

b) $N > 1$ and a $N \times N$ permutation matrix $\underline{\underline{P}}$ exists such that

$$\underline{\underline{P}}^T \underline{\underline{A}}\, \underline{\underline{P}} = \begin{array}{c} \left[\begin{array}{c|c} \underline{\underline{B}} & \underline{\underline{C}} \\ \hline \underline{\underline{0}} & \underline{\underline{D}} \end{array}\right] \begin{array}{l} \}\, r \\ \}\, N-r \end{array} \\ \begin{array}{cc} \underbrace{\quad}_{r} & \underbrace{\quad}_{N-r} \end{array} \end{array}$$

this implies that some of the unknowns may be decoupled

and an $N \times N$ matrix is called irreducible if it is not reducible.

Thm If $\underline{\underline{A}}$ is irreducible and diagonally dominant and if

$$\bigstar \quad |a_{ii}| > \sum_{\substack{j=1 \\ j \ne i}}^{N} |a_{ij}|$$

for at least one value for i (strictly diagonally dominant for just one i)

then $\underline{\underline{A}}$ is nonsingular.

For the discretization with $c = 0$, we have that $\underline{\underline{A}}$ is irreducible (because all equations are coupled) and diagonally dominant and $\bigstar$ holds for grid points near the boundary $\Rightarrow$ the system is solvable.

When $c < 0$ it is possible that c is an eigenvalue of the problem

$$\left(\frac{-1}{\Delta x^2} \delta_x^2 - \frac{1}{\Delta y^2} \delta_y^2 + \frac{a}{2\Delta x} \delta_{0x} + \frac{b}{2\Delta y} \delta_{0y} \right) v_{j,k} = \lambda v_{j,k}$$

so that the matrix would become singular.

Convergence

Let $Lu = -u_{xx} - u_{yy}$

Poisson equation: $Lu = f(x,y)$, $0<x<1$, $0<y<1$

$u = g(x,y)$ on $\partial\Omega$

Finite difference approximation

$$L_h v_h = \frac{-1}{\Delta x^2}\delta_x^2 v_{j,k} - \frac{1}{\Delta y^2}\delta_y^2 v_{j,k} = f(x_j, y_k) \qquad \begin{array}{l} 1 \le j \le N-1 \\ 1 \le k \le M-1 \end{array}$$

let G_0 be the interior of grid, ie $\{1 \le j \le N-1,\ 1 \le k \le M-1\}$

boundary conditions: $v_{j,k} = g(x_j, y_k)$ on ∂G

Problem: show convergence, ie

$$\max_G |v_{j,k} - u(x_j, y_k)| \to 0 \quad \text{as} \quad \Delta x, \Delta y \to 0$$

First consider the truncation error

$$\tau_{j,k} = f(x_j, y_k) - L_h u(x_j, y_k) = O(\Delta x^2, \Delta y^2)$$

by the usual Taylor series method.

Therefore, $\max_G |\tau_{j,k}| \to 0$ as $\Delta x, \Delta y \to 0$

and the scheme is consistent.

To go from consistency to convergence, require some form of regularity of L_h. (Like stability for a time-dependent problem, but the analysis is different.)

Notation

$$\|v_h\|_G = \max_{(j,k)\in G} |v_{j,k}| \qquad \text{(on the entire grid)} \quad 0 \le j \le N,\ 0 \le k \le M$$

$$\|v_h\|_{G_0} = \max_{(j,k)\in G_0} |v_{j,k}| \qquad \text{(on the interior of grid)} \quad 1 \le j \le N-1,\ 1 \le k \le M-1$$

$$\|v_h\|_{\partial G} = \max_{(j,k)\in \partial G} |v_{j,k}| \qquad \text{(on the boundary only)} \quad j = 0, N \quad k = 0, M$$

Discrete max/min principle

Theorem: If $L_h v_h \le 0$ ($L_h v_h \ge 0$) on G_0, then the maximum (minimum) value of $v_{j,k}$ on G occurs on ∂G.

proof: Note that $L_h v_h \le 0$ implies

$$-\frac{1}{\Delta x^2}\delta_x^2 v_{j,k} - \frac{1}{\Delta y^2}\delta_y^2 v_{j,k} \le 0$$

$$\left(\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}\right)v_{j,k} \le \frac{1}{2}\left[\frac{1}{\Delta x^2}\left(v_{j+1,k}+v_{j-1,k}\right)+\frac{1}{\Delta y^2}\left(v_{j,k+1}+v_{j,k-1}\right)\right]$$

Suppose that $v_{j,k}$ is a local max on G_0:

$v_{j,k} \ge v_{j+1,k}$ $\quad$ $v_{j,k} \ge v_{j-1,k}$

$v_{j,k} \ge v_{j,k+1}$ $\quad$ $v_{j,k} \ge v_{j,k-1}$

use these in the formula above

$$\left(\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}\right)v_{j,k} \le \frac{1}{2}\left[\frac{2}{\Delta x^2}v_{j,k}+\frac{1}{\Delta y^2}\left(v_{j,k}+v_{j,k-1}\right)\right]$$

if you use all four inequalities

$$\left(\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}\right)v_{j,k} \le \underbrace{\frac{1}{2}\left[\frac{1}{\Delta x^2}\left(v_{j,k}+v_{j,k}\right)+\frac{1}{\Delta y^2}\left(v_{j,k}+v_{j,k}\right)\right]}_{\text{since this quantity is bounded above and below, the inequalities are equalities}} \le \left(\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}\right)v_{j,k}$$

Then $v_{j,k} = v_{j+1,k} = v_{j-1,k} = v_{j,k-1} = v_{j,k+1}$

and therefore $v_{j,k}$ is not a local max.

We now use the discrete max/min principle to prove a regularity result.

<u>Theorem</u>: Suppose that $w_{j,k}$ is a grid function defined on G with $w_{j,k} = 0$ on ∂G, then $\| w \|_G \le \frac{1}{8} \| L_h w_{j,k} \|_{G_0}$

<u>Proof</u>: Define $F_{j,k} = L_h w_{j,k}$

and note that $-\| F \|_{G_0} \le L_h w_{j,k} \le \| F \|_{G_0}$

Define $z_{j,k} = \frac{1}{4}\left[(x_j - \frac{1}{2})^2 + (y_k - \frac{1}{2})^2 \right]$ on G

Note that $L_h z_h = -1$

To see this, $\delta_x^2 (x_j - \frac{1}{2})^2 = (x_{j-1} - 1/2)^2 - 2(x_j - 1/2)^2 + (x_{j+1} - 1/2)^2$

$= (x_j - \Delta x - \frac{1}{2})^2 - 2(x_j - \frac{1}{2})^2 + (x_j + \Delta x - \frac{1}{2})^2$

$= (x_j - \Delta x)^2 - (x_j - \Delta x) + \frac{1}{4} - 2(x_j - \frac{1}{2})^2 + (x_j + \Delta x)^2 - (x_j + \Delta x) + \frac{1}{4}$

$= \cancel{x_j^2} - \cancel{2\Delta x x_j} + \Delta x^2 - \cancel{x_j} + \cancel{\Delta x} + \cancel{\frac{1}{4}} - 2(\cancel{x_j^2} - \cancel{x_j} + \cancel{\frac{1}{4}}) + \cancel{x_j^2} + \cancel{2 x_j \Delta x} + \Delta x^2 - \cancel{x_j} - \cancel{\Delta x} - \cancel{\frac{1}{4}}$

$= 2\Delta x^2$

$\rightarrow L_h z_h = \frac{-1}{4}\left[\frac{2\Delta x^2}{\Delta x^2} + \frac{2\Delta y^2}{\Delta y^2} \right] = -1$

Next, $L_h \left(w_h - \| F \|_{G_0} z_h \right) = L_h w_h + \| F \|_{G_0} \ge 0$

$L_h \left(w_h + \| F \|_{G_0} z_h \right) = L_h w_h - \| F \|_{G_0} \le 0$

Now by the discrete max/min principle

since $L_h \left(w_h - \| F \|_{G_0} z_h \right) \ge 0$ then

$\min \left(w_h - \| F \|_{G_0} z_h \right)$ occurs on boundary

and since $L_h \left(w_h + \| F \|_{G_0} z_h \right) \le 0$ then

$\max \left(w_h + \| F \|_{G_0} z_h \right)$ occurs on boundary

So we can write

$$\max_G \left(w_h + \|F\|_{G_0} z_h\right) = \max_{\partial G} \left(w_h + \|F\|_{G_0} z_h\right) = \|F\|_{G_0} \max_{\partial G} z_h$$

because we defined $w_{j,k} = 0$ on ∂G

$$\Rightarrow \quad w_h \leq \|F\|_{G_0} \|z_h\|_{\partial G}$$

likewise, $w_h \geq -\|F\|_{G_0} \|z_h\|_{\partial G}$ where $\|z_h\|_{\partial G} = \frac{1}{8}$

Therefore we can write

$$-\frac{1}{8}\|L_h w_h\|_{G_0} \leq w_h \leq \frac{1}{8}\|L_h w_h\|_{G_0}$$

END OF PROOF

<u>Back to convergence:</u>

note, $w_{j,k} = 0$ on boundary due to Dirichlet conditions

set $w_{j,k} = v_{j,k} - u(x_j, y_k)$

Then $L_h w_h = L_h v_h - L_h u(x_j, y_k) = \tau_{j,k}$

The regularity result gives us

$$\|w_h\|_G \leq \frac{1}{8}\|L_h w_h\|_{G_0} = \frac{1}{8}\|\tau_h\|_{G_0} = O(\Delta x^2, \Delta y^2)$$

$$\rightarrow \quad \boxed{\|w_h\|_G \leq O(\Delta x^2, \Delta y^2) \quad \text{as } \Delta x, \Delta y \to 0}$$

Solution Schemes

Model problem:

$$\frac{1}{\Delta x^2}\,\delta_x^2 v_{j,k} + \frac{1}{\Delta y^2}\,\delta_y^2 v_{j,k} = F(x_j, y_k) \qquad \begin{array}{l} 1 \le j \le N-1 \\ 1 \le k \le M-1 \end{array}$$

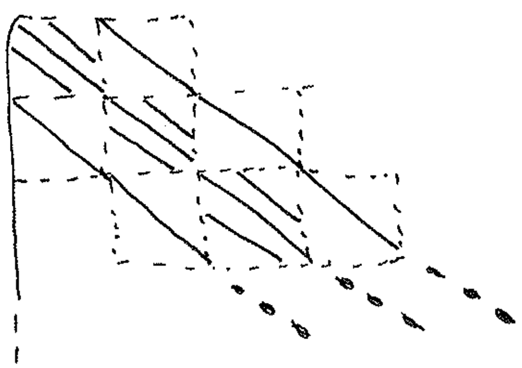
with $v_{j,k} = g(x_j, y_k)$ for j, k on ∂G

This implies a linear system $\underline{\underline{A}}\,\underline{v} = \underline{c}$

where $\underline{\underline{A}}$ was block tri-diagonal

The problem boils down to linear algebra and whether you employ direct or iterative methods.

Direct Methods

$\underline{\underline{A}} =$ 

The diagonal matrices are tridiagonal and the sub and super diagonal matrices are diagonal

the matrix $\underline{\underline{A}}$ is banded with bandwidth $= 2\min(N-1, M-1) + 1$

ie bandwith is $O(\min(N,M))$

Use a direct banded matrix solver based on Gaussian elimination with partial pivoting: this results in an operation count $= O(\min(N,M)^2 \cdot NM)$

if ~~ø~~ $N = M$ then operation count $= O(N^4)$

$$\boxed{\text{this is a safe (reliable, stable) method but slow}}$$

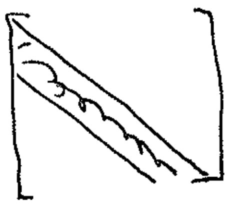
The optimal operation count is $O(N^2)$, ie one calculation for every grid point.

Direct Factorization

Since $\underline{\underline{A}}$ is symmetric, positive definite, you could use Choleskyi.

$$\underline{\underline{A}} = \underline{\underline{L}}\,\underline{\underline{L}}^T$$

where $\underline{\underline{L}}$ is lower triangular and retains the same band structure as original matrix $\underline{\underline{A}}$

$\underline{\underline{L}} =$  the bandwith is $\frac{1}{2}$ the orignal bandwith

operation count $= O(N^4)$

Block Tridiagonal Solvers

$$\underline{\underline{A}} = \begin{bmatrix} C_1 & D_1 & & & \\ B_2 & C_2 & D_2 & & \\ & B_3 & C_3 & D_3 & \ddots \\ & & \ddots & \ddots & \ddots \\ & & & & \end{bmatrix}$$

where B, C, D are $(N-1)\times(N-1)$ matrices

write down the augmented matrix

$$\left[\begin{array}{cccc|c} C_1 & D_1 & & & F_1 \\ B_2 & C_2 & D_2 & & F_2 \\ & \ddots & \ddots & \ddots & \\ & & & & \\ & & & & \end{array}\right]$$

$\leftarrow$ multiply by $-B_2 C_1^{-1}$ and add to second row

$$\rightarrow \quad \tilde{C}_2 = C_2 - B_2 C_1^{-1} D_1$$

$$\tilde{F}_2 = F_2 - B_2 C_1^{-1} F_1$$

let $C_1^{-1} D_1 = Z_1 \rightarrow C_1 Z_1 = D_1$

where C_1 is tridiagonal

cost to compute Z_1 is $O(N^2)$

ultimately, using the tridiagonal block solvers, the operational cost remains $O(N^4)$

Iterative Methods (Residual Correction Methods)

Let $\underline{w}$ denote an approximation of $\underline{v}$.

Then $\boxed{\begin{aligned} \underline{e} &= \underline{w} - \underline{v} = \text{error} \\ \underline{r} &= \underline{c} - \underline{\underline{A}}\,\underline{w} = \text{residual} \end{aligned}}$ and

The error and residual are related

$$\underline{\underline{A}}\,\underline{e} = \underline{\underline{A}}(\underline{w} - \underline{v}) = \underline{\underline{A}}\,\underline{w} - \underline{c} = -\underline{r} \implies \boxed{\underline{e} = -\underline{\underline{A}}^{-1}\underline{r}}$$

Then the solution

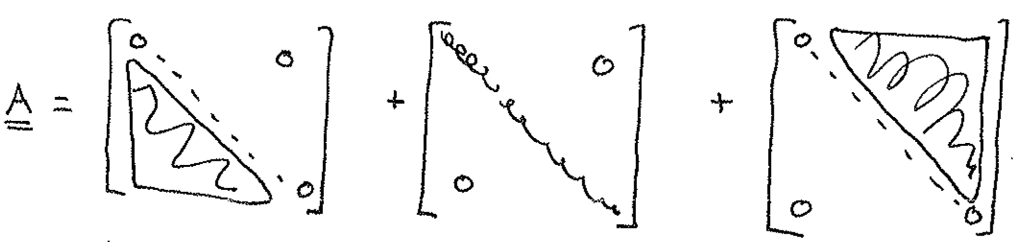
$$\underline{v} = \underline{w} - \underline{e} = \underline{w} + \underline{\underline{A}}^{-1}\underline{r} \longrightarrow \boxed{\underline{v} = \underline{w} + \underline{\underline{A}}^{-1}\underline{r}}$$

we don't have $\underline{\underline{A}}^{-1}$ (it is expensive to calculate).
However, suppose that $\underline{\underline{B}}$ approximates $\underline{\underline{A}}^{-1}$, then

$$\boxed{\underline{w}^{n+1} = \underline{w}^{n} + \underline{\underline{B}}\,\underline{r}^{n} \quad \text{"residual" correction scheme}}$$

There are several choices of $\underline{\underline{B}}$:

Let $\underline{\underline{A}} = \underline{\underline{L}} + \underline{\underline{D}} + \underline{\underline{U}}$ where



$\underline{\underline{L}}$ lower triangular portion

$\underline{\underline{D}}$ diagonal portion

$\underline{\underline{U}}$ upper triangular portion

Choices for $\underline{\underline{B}}$

① Jacobi choice: $\underline{\underline{B}} = \underline{\underline{D}}^{-1}$

② Gauss-Seidel $\underline{\underline{B}} = (\underline{\underline{L}} + \underline{\underline{D}})^{-1}$

③ Successive Over Relaxation (SOR)

$\underline{\underline{B}} = \omega(\underline{\underline{D}} + \omega\underline{\underline{L}})^{-1}$ where ω is a relaxation parameter (ω is a scalar)

Comments:

- the implementation is very simple
- convergence occurs because $\underline{\underline{A}}$ comes from the discretization of an elliptic PDE

11/24/03

Example: consider a 1D elliptic problem, $-u_{xx} = f$, $u=0$ on $\partial\Omega$

$$-\frac{1}{\Delta x^2}\delta_x^2 v_j = f_j, \quad 1 \le j \le N-1, \quad v_0 = v_N = 0$$

Suppose $N=4$

$$\begin{cases} -v_2 + 2v_1 = \Delta x^2 f_1 \\ -v_3 + 2v_2 - v_1 = \Delta x^2 f_2 \\ 2v_3 - v_2 = \Delta x^2 f_3 \end{cases}$$

this is a 3x3 linear system

$\underline{\underline{A}}\,\underline{v} = c \rightarrow \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \Delta x^2 \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$

Suppose we instead "solve" for the diagonal elements:

$$v_1 = \frac{1}{2}(\Delta x^2 f_1 + v_2)$$

$$v_2 = \frac{1}{2}(\Delta x^2 f_2 + v_1 + v_3)$$

$$v_3 = \frac{1}{2}(\Delta x^2 f_3 + v_2)$$

add and subtract diagonal

$$
\begin{aligned}
v_1 &= v_1 + \tfrac{1}{2}\left(\Delta x^2 f_1 + v_2 - 2v_1\right)\\
v_2 &= v_2 + \tfrac{1}{2}\left(\Delta x^2 f_2 + v_1 - 2v_2 + v_3\right)\\
v_3 &= v_3 + \tfrac{1}{2}\left(\Delta x^2 f_3 + v_2 - 2v_3\right)
\end{aligned}
$$

this suggests some sort of iterative scheme:

$$
\left.
\begin{aligned}
v_1^{n+1} &= v_1^n + \tfrac{1}{2}\left(\Delta x^2 f_1 + v_2^n - 2v_1^n\right)\\
v_2^{n+1} &= v_2^n + \tfrac{1}{2}\left(\Delta x^2 f_2 + v_1^n - 2v_2^n + v_3^n\right)\\
v_3^{n+1} &= v_3^n + \tfrac{1}{2}\underbrace{\left(\Delta x^2 f_3 + v_2^n - 2v_3^n\right)}_{\text{residual}}
\end{aligned}
\right\}\ \underline{\text{Jacobi}}
$$

If use the n+1 term right away:

$$
\left.
\begin{aligned}
v_1^{n+1} &= v_1^n + \tfrac{1}{2}\left(\Delta x^2 f_1 + v_2^n - 2v_1^n\right)\\
v_2^{n+1} &= v_2^n + \tfrac{1}{2}\left(\Delta x^2 f_2 + v_1^{n+1} - 2v_2^n + v_3^n\right)\\
v_3^{n+1} &= v_3^n + \tfrac{1}{2}\underbrace{\left(\Delta x^2 f_3 + v_2^{n+1} - 2v_3^n\right)}_{\text{residual}}
\end{aligned}
\right\}\ \underline{\text{Gauss - Seidel}}
$$

Include a relaxation parameter ω

$$
\left.
\begin{aligned}
v_1^{n+1} &= v_1^n + \tfrac{\omega}{2}\left(\Delta x^2 f_1 + v_2^n - 2v_1^n\right)\\
v_2^{n+1} &= v_2^n + \tfrac{\omega}{2}\left(\Delta x^2 f_2 + v_1^{n+1} - 2v_2^n + v_3^n\right)\\
v_3^{n+1} &= v_3^n + \tfrac{\omega}{2}\left(\Delta x^2 f_3 + v_2^{n+1} - 2v_3^n\right)
\end{aligned}
\right\}\ \underline{\text{SOR}}
$$

Typical Algorithm

$$\left(\frac{-1}{\Delta x^2}\delta_x^2 - \frac{1}{\Delta y^2}\delta_y^2\right) v_{j,k} = f_{j,k} \quad , \quad 1 \le j \le N-1 , \quad 1 \le k \le M-1$$

with $v_{j,k} = 0$ on boundary

For SOR:

1) Set $v_{j,k} = 0$ everywhere, pick ω
2) For $n = 1, 2, \ldots, n_{max}$
3) for $k = 1, \ldots, M-1$
4) for $j = 1, \ldots, N-1$
5) $v_{j,k} = v_{j,k} + \frac{\omega}{D}\left(f_{j,k} + \frac{1}{\Delta x^2}\delta_x^2 v_{j,k} + \frac{1}{\Delta y^2}\delta_y^2 v_{j,k}\right)$ $\quad D = \frac{2}{\Delta x^2} + \frac{2}{\Delta y^2}$
6) stop if $\max|r_{j,k}| < tol$

one sweep (steps 3–5)

cost per sweep: $O(MN)$

total cost to converge: $O(\underbrace{n}_{\text{number of iterations}} MN)$

so the question is, how many iterations are necessary for convergence? $n \sim O(MN)$

so we're back to $O(N^4)$ - these iterative methods haven't bought us much over the direct methods. Need to use multigrid...

Analysis of Residual Correction Schemes

Iteration: $\underline{w}^{n+1} = \underline{w}^n + \underline{\underline{B}}\,\underline{r}^n$

error eqn: $\underline{\underline{A}}\,\underline{e}^n = -\underline{r}^n$ where $\underline{e}^n = \underline{w}^n - \underline{v}$

eliminate $\underline{r}^n$: $\underline{w}^{n+1} = \underline{w}^n - \underline{\underline{B}}\,\underline{\underline{A}}\,\underline{e}^n$

subtract $\underline{v}$ from both sides

$$\underline{e}^{n+1} = \underline{e}^n - \underline{\underline{B}}\,\underline{\underline{A}}\,\underline{e}^n$$

$$\underline{e}^{n+1} = (\underline{\underline{I}} - \underline{\underline{B}}\,\underline{\underline{A}})\,\underline{e}^n$$

Suppose $\underline{e}^0$ is the initial error, then

power, not time step

$$\underline{e}^n = \underbrace{\left[\underline{\underline{I}} - \underline{\underline{B}}\,\underline{\underline{A}}\right]}_{\underline{\underline{R}}}{}^{n}\,\underline{e}^0$$

the iteration converges if

power

$$\lim_{n\to\infty} \underline{\underline{R}}^n\,\underline{e}^0 = 0$$

<u>Theorem</u>: convergence occurs for any $\underline{e}^0$ if $\underline{\underline{R}}$ is a convergent ~~radius~~ matrix, ie <u>iff</u> spectral radius of R less than one.

So the rate of convergence depends on the eigenvalues of $\underline{\underline{R}}$. Suppose $\underline{\underline{R}}$ has eigenvalues

$$|\lambda_1| > |\lambda_2| > |\lambda_3| > \dots$$

dominant eigenvalue (controlls convergence)

analysis of residual corr schemes

11/24/03

80

Define $\{\alpha_j\}_{j=1}^{N}$ such that

$$\underline{e}^{\circ} = \sum_{j=1}^{N} \alpha_j \underline{\xi}_j$$ where $\underline{\xi}_j$ ~~each~~ is the j^{th} eigenvector of $\underline{\underline{R}}$, which is $N \times N$ matrix

$$\rightarrow \quad \underline{e}^{1} = \underline{\underline{R}}\, \underline{e}_0 = \sum_{j=1}^{N} \alpha_j \underline{\underline{R}}\, \underline{\xi}_j = \sum_{j=1}^{N} \alpha_j \lambda_j \underline{\xi}_j$$

$$\rightarrow \quad \underline{e}^{2} = \sum_{j=1}^{N} \alpha_j \lambda_j^2 \underline{\xi}_j$$

$$\vdots$$

$$\underline{e}^{n} = \sum_{j=1}^{N} \alpha_j \lambda_j^n \underline{\xi}_j$$

$$\underline{e}^{n} = \lambda_1^n \left(\alpha_1 \underline{\xi}_1 + \underbrace{\sum_{n=1}^{N} \alpha_j \left(\frac{\lambda_j}{\lambda_1}\right)^n \underline{\xi}_j}_{\to 0} \right)$$

$\rho(\underline{\underline{R}})$ - spectral radius of $\underline{\underline{R}}$ - magnitude of largest λ_j

$$\rightarrow \quad \underline{e}^n \sim \lambda_1^n \alpha_1 \underline{\xi}_1 \to 0$$

$$\Rightarrow \quad \frac{\| \underline{e}^{n+1} \|}{\| \underline{e}^{n} \|} \sim |\lambda_1| \quad \rightarrow \quad \frac{\| \underline{e}^{n} \|}{\| \underline{e}^{\circ} \|} \sim |\lambda_1|^n < \text{tol}$$

$$\Rightarrow \quad n \log(\rho(\underline{\underline{R}})) < \log(\text{tol}) \quad \rightarrow \quad \boxed{n > \frac{\log(\text{tol})}{\log(\rho(\underline{\underline{R}}))} = O\left(\frac{-1}{\log(\rho(\underline{\underline{R}}))}\right)}$$

IF $\rho(\underline{\underline{R}}) \sim 0 \Rightarrow$ fast convergence

IF $\rho(\underline{\underline{R}}) \sim 1-\epsilon \Rightarrow$ slow convergence

Suppose $\quad -\delta_x^2 v_j = \Delta x^2 f_j \quad , \quad 1 \le j \le N-1$

$v_0, v_N = 0$

This leads to a system $\underline{\underline{A}}\,\underline{v} = \underline{c}$ whose eigenvalues are $\nu_p = 2\left(1 - \cos\left(\frac{p\pi}{N}\right)\right)$, $p = 1, \dots, N-1$.

Note - these are eigenvalues of $\underline{\underline{A}}$, not $\underline{\underline{R}}$, but they are related.

$\underline{\underline{D}} = 2\,\underline{\underline{I}}$

For the Jacobi method, $\underline{\underline{R}} = \underline{\underline{I}} - \underline{\underline{D}}^{-1}\underline{\underline{A}}$

$\rightarrow \underline{\underline{R}} = \underline{\underline{I}} - \frac{1}{2}\underline{\underline{A}}$

eigenvalue problem for $\underline{\underline{R}}$: $\underline{\underline{R}}\,\underline{\xi} = \lambda\,\underline{\xi}$

$\left(\underline{\underline{I}} - \frac{1}{2}\underline{\underline{A}}\right)\underline{\xi} = \lambda\,\underline{\xi} \quad \longrightarrow \quad -\frac{1}{2}\underline{\underline{A}}\,\underline{\xi} = (\lambda - 1)\,\underline{\xi}$

$\underline{\underline{A}}\,\underline{\xi} = \underbrace{-2(\lambda-1)}_{\nu}\,\underline{\xi} \quad \longrightarrow \quad \nu = -2(\lambda - 1) \quad \longrightarrow \quad \lambda = 1 - \frac{\nu}{2}$

$\longrightarrow \quad \lambda_p = \cos\left(\frac{p\pi}{N}\right) \quad \Rightarrow \quad \rho(\underline{\underline{R}}) = \cos\frac{\pi}{N} = 1 - \frac{1}{2}\left(\frac{\pi}{N}\right)^2 + \dots$

really interested in logarithm of spectral radius

$\ln\left(\rho(\underline{\underline{R}})\right) \approx -\frac{1}{2}\left(\frac{\pi}{N}\right)^2 + \dots$

$\longrightarrow \quad n_{\text{iterate}} \sim \frac{-1}{\ln(\rho(\underline{\underline{R}}))} \sim C N^2$

Multigrid Methods

Multigrid is a technique to accelerate the convergence of a residual correction method.

Elliptic (Linear) PDE : $Lu = f$, $\underline{x} \in \Omega$

$u = g$, $\underline{x} \in \partial\Omega$

Finite difference discretization: $L_h v_h = f_h$, $\underline{x} \in \Omega_h$

$v_h = g_h$, $\underline{x} \in \partial\Omega_h$

assume that $v_h \in \mathbb{R}^n$ - (v_h is a grid function in $\mathbb{R}^n$)

test problem: $Lu = u_{xx} = f(x)$, $0 \le x \le 1$

$u(0) = u(1) = 0$

approximation: $v_{j+1} - 2v_j + v_{j-1} = h^2 f(x_j)$, $1 \le j \le N-1$

$v_0 = v_N = 0$

residual correction method based on Jacobi

$$v_j^{n+1} = v_j^n + \frac{\omega}{2}\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n - h^2 f(x_j)\right)$$ - (ω - Jacobi) (NOT SOR)

if $\omega = 1$, this is classical Jacobi
if $0 < \omega < 1$, under-relaxed Jacobi } converges take $v_j^0 = 0$
if $\omega > 1$, over-relaxed Jacobi } diverges

in matrix form

$$\underline{v}^{n+1} = \left(\underline{\underline{I}} + \frac{\omega}{2}\underline{\underline{A}}\right)\underline{v}^n - \frac{\omega h^2}{2}\underline{f}$$

$$\underline{v}^n = \begin{bmatrix} v_1^n \\ \vdots \\ v_{N-1}^n \end{bmatrix}, \quad \underline{\underline{A}} = \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \end{bmatrix}, \quad \underline{f} = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_{N-1}) \end{pmatrix}$$

Error in the $n^{\underline{th}}$ iterate

$$\underline{e}^n = \underline{v}^n - \underline{v} \qquad (\underline{v} \text{ is the true steady-state error})$$

by a similar procedure as just done, you find

$$\underline{e}^{n+1} = \left(\underline{\underline{I}} + \frac{w}{2} \underline{\underline{A}} \right) \underline{e}^n = \underline{\underline{R}}\, \underline{e}^n$$

turns into eigenvalue problem,

$$\underline{\underline{R}}\, \underline{\xi} = \lambda \underline{\xi}$$

write $\underline{e}^0$ as an eigenvalue expansion

$$\underline{e}^0 = \alpha_1 \underline{\xi}_1 + \alpha_2 \underline{\xi}_2 + \cdots + \alpha_{N-1} \underline{\xi}_{N-1}$$

where $\underline{\xi}_j$ is the $j^{\underline{th}}$ eigenvector of $\underline{\underline{I}} + \frac{w}{2} \underline{\underline{A}}$

and whose eigenvalues are λ_j

From before, $\boxed{\underline{e}^n = \alpha_1 \lambda_1^n \underline{\xi}_1 + \alpha_2 \lambda_2^n \underline{\xi}_2 + \cdots}$

<u>we want to study the behavior of this expansion</u>

Have $\lambda_p = 1 - w\left(1 - \cos\left(\frac{p\pi}{N}\right)\right)$ - eigenvalues of $\underline{\underline{R}}$, $p = 1, \ldots, N-1$

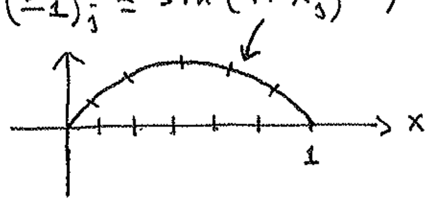
j component of p^{th} eigenvector $(\underline{\xi}_p)_j = \sin\left(\frac{jp\pi}{N}\right)$ - the j^{th} component of $p^{\underline{th}}$ eigenvector

<u>we examine the "modes" of the eigenvector expansion</u>

Consider a specific case, $N = 6$

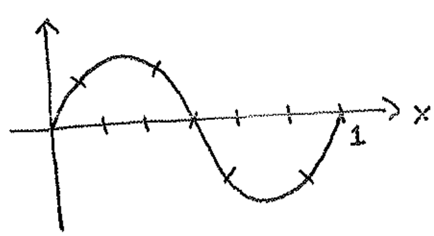
$$(\underline{\xi}_p)_j = \sin\left(\frac{jp\pi}{6}\right) = \sin(p\pi x_j), \quad x_j = \frac{j}{6}$$

- $p = 1$: $(\underline{\xi}_1)_j = \sin(\pi x_j)$, $\lambda_1 = 1 - w\left(1 - \frac{\sqrt{3}}{2}\right)$

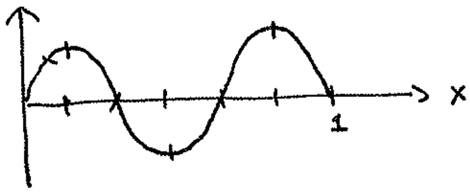


slowest mode to converge

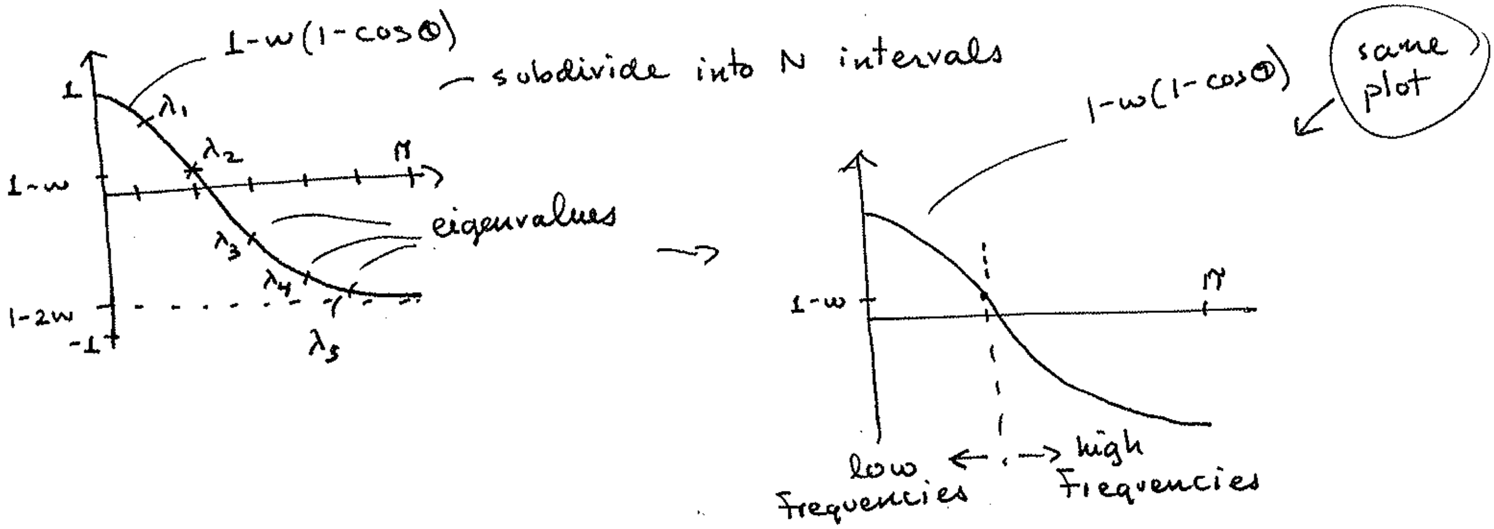
- $p = 2$ $(\underline{\xi}_2)_j = \sin(2\pi x_j)$, $\lambda_2 = 1 - \frac{w}{2}$



$p = 3$, $\sin(3\pi x_j)$, $\lambda_3 = 1 - \omega$



the point - the frequency represented by eigenvectors increases with p

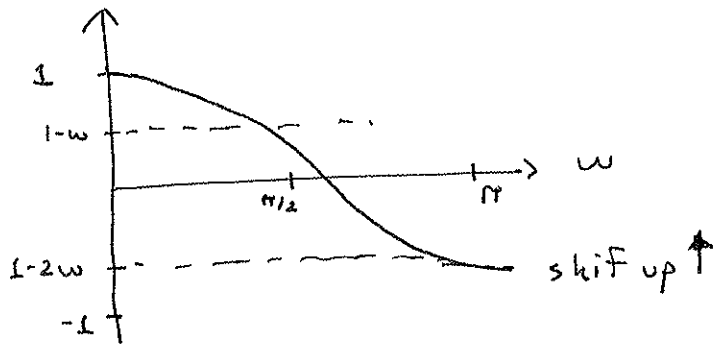


Key observations that lead us to Multigrid

1) Convergence occurs for $0 < \omega < 1$
2) If you are in range of convergence, $0 < \omega < 1$ what is spectral radius? $\rho(R) = \lambda_1 = 1 - \omega(1 - \cos\frac{\pi}{N})$
$\sim 1 - \frac{K}{N^2} \Rightarrow$ slow convergence for any choice of ω

If convergence is slow, why use ω at all?

3) Low frequency modes always converge slowly but the high frequency modes can be made to converge quickly with a good choice of ω.



Choose ω such that
$$(1-\omega) = -(1-2\omega)$$
$$\Rightarrow \quad \omega = \frac{2}{3}$$

when $\omega = \frac{2}{3}$, the spectral radius only for the high frequency modes: $\rho(R)_{\substack{high \\ Freq}} = \frac{1}{3}$ - bounded away from 1 independently ~~from~~ of N.

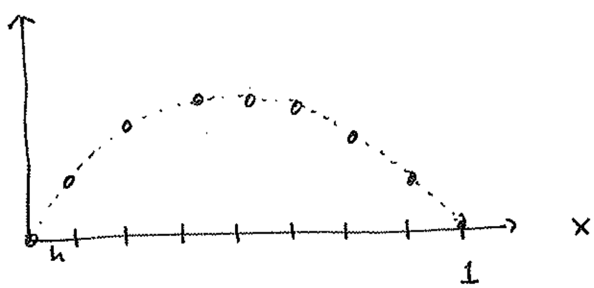
The multigrid idea is to use coarser grids to resolve low frequency components of the error.

Two-Grid Algorithm

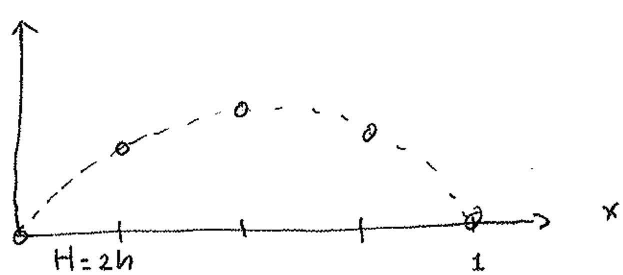
Use grids with $h = \frac{1}{N}$ (Fine grid), ultimately the grid on which we want the solution.

and use grids with $H = 2h$ (coarse grid)

Fine grid ($N=8$)



coarse grid ($N=4$)



Two-Grid Algorithm (continued)

step 0 : initial guess, $\underline{v}^0$ given on fine grid

step ① : Apply ν_1 steps of ω-Jacobi (smoothing step)

$$\underline{v}^k = (\underline{\underline{I}} + \tfrac{\omega}{2}\underline{\underline{A}})\,\underline{v}^{k-1} - \tfrac{\omega}{2} h^2 \underline{f} \qquad k = 1, \dots, \nu$$

$\tilde{\underline{v}} = \underline{v}^{\nu_1}$, the new vector after ν_1 steps of ω-Jacobi

the number ν_1 is chosen adaptively, but for simple code, pick $\nu_1 \approx 3$ or 4 - a small number then $\tilde{\underline{v}}$ only has low frequency components of error

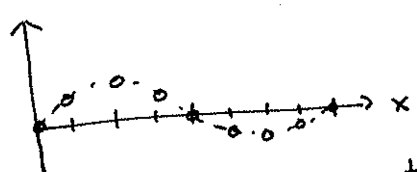
step ② : compute residual and restrict to the coarse grid

$$\underline{r} = \underline{f} - \underline{\underline{A}}\,\tilde{\underline{v}} \quad \text{(smooth}$$

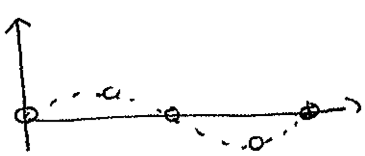
then

$$\underline{r}_H = \underline{\underline{Q}}\, r$$

suppose this is residual on fine grid



simply take every other point for the coarse grid.



then for this choice, $\underline{\underline{Q}}$ is (not square)

$$\begin{bmatrix} \underline{r}_H \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \end{bmatrix}}_{\underline{\underline{Q}}} \begin{bmatrix} \underline{r} \end{bmatrix}$$

instead of creating matrix and operating on $\underline{r}$, just pick out the values on the coarse grid for $\underline{r}_H$

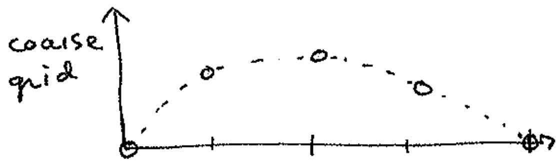
<u>step 3</u> Solve the error equation on coarse grid

$$\underline{\underline{A}}_H \, \underline{e}_H = \underline{r}_H$$

(this is equivalent to solving the discrete problem on coarse grid)

<u>step 4</u> interpolate $\underline{e}_H$ on to fine grid error, $\underline{e}$

$$\underline{e} = \underline{\underline{P}} \, \underline{e}_H$$



you could use piecewise linear interpolation of coarse grid values to find fine grid values

you don't necessarily need fancy interpolation because error found in high frequency can be smoothed out by a few Jacobi iterates

$$\begin{bmatrix} \\ \underline{e} \\ \\ \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1/2 & 1/2 & \cdots & \\ 0 & 1 & \cdots & \\ & & & \end{bmatrix}}_{\underline{\underline{P}}} \begin{bmatrix} \underline{e}_H \end{bmatrix}$$

<u>step 5</u> update and smooth

$$\bar{\underline{v}} = \tilde{\underline{v}} + \underline{e}$$ ← here we have filtered out the low frequency error in $\tilde{v}$ by adding $\underline{e}$; we've introduced lower frequency error due to interpolation, but we now smooth that out

set $\underline{v}^0 = \bar{\underline{v}}$

$$\underline{v}^k = \left(\underline{\underline{I}} + \frac{\omega}{2} \underline{\underline{A}} \right) \underline{v}^{k-1} - \frac{\omega}{2} h^2 \underline{F}, \quad k = 1, \ldots, \nu_2$$

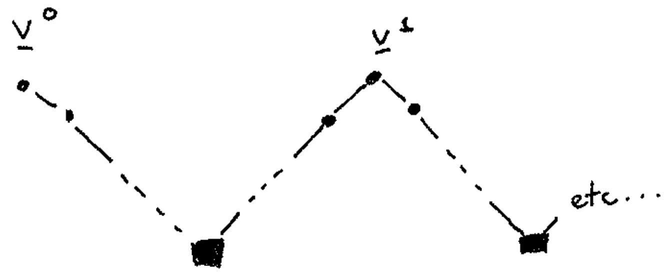
then $\hat{\underline{v}} = v^{\nu_2}$ ← this is the next iterate

at this point we've taken one full two-grid algorithm iteration

Multigrids

levels

fine k
$k-1$
$\vdots$
coarse 1



this is known as "v" cycle

each • represents a smoothing operation

↘ a down error represents the restriction onto a coarser grid

↗ an up error represents an interpolation

■ represents a solution solve

total operation count: $\propto$ # grid points

Final Exam

Part 1 - in class, thursday

OPEN NOTES, CLOSED BOOKS

$\approx 75\%$ of final exam

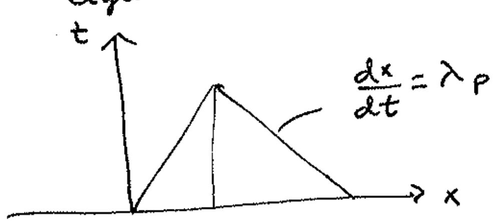
Part 2: take home, available starting saturday morning

$\approx 25\%$ email for copy, 48 hrs to complete

Topics since last midterm

1) Linear hyperbolic PDEs, $\underline{u}_t + \underline{\underline{A}}\,\underline{u}_x = 0$

hyperbolic if A is diagonalizable with real eigenvalues



Basic schemes

Lax-Friedrichs 1st order

Lax-Wendroff 2nd order

upwind methods 1st order

both are considered centered methods

one-sided scheme

behavior near discontinuities
(dissipative versus dispersive)
determined by examining modified equation
stability and CFL condition
numerical vs exact domain of dependence
non-constant coeff linear PDEs
boundary conditions (inflow vs outflow)

2) Hyperbolic conservation Laws

$u_t + f(u)_x = 0$

integral form (integral conservation form)

$$\frac{d}{dt}\int_a^b u\,dx = -f(u)\Big|_a^b$$

shock conditions, method of characteristics

Riemann problems

numerical methods

- conservative finite volume schemes
- numerical flux functions (for LF, LW, etc...)
- Lax-Wendroff theorem
- Godunov's method (nonlinear upwind method, 1st order accurate)
- high resolution methods (2nd order at least, but 1st order near shocks)
  - flux limiters, slope limiters

3) Elliptic PDEs

canonical forms

properties of Laplace's eqn

- mean value, maximum principle

finite difference methods

discretizations, solvability of $Ax=b$, convergence

solution schemes (direct vs iterative methods)

multigrid