2016 02 22

Partial Differential Equations

Computational methods for PDEs:
- Finite difference methods
- Parabolic PDEs

A second order PDE:

$$x,y\in\Omega \quad A(x,y)\partial_{xx}u + B(x,y)\partial_{xy}u + C(x,y)\partial_{yy}u = w(u,\partial_x u,\partial_y u, x, y)$$

- Parabolic if $\forall x,y\in\Omega \ B^2-4AC=0$
- Hyperbolic if $\forall x,y\in\Omega \ B^2-4AC>0$
- Elliptic if $\forall x,y\in\Omega \ B^2-4AC<0$

- Mostly Linear PDEs

$$\partial_t u = \partial_x(a(x,t)\partial_x u) + b(x,t)\partial_x u + c(x,t)u + d(x,t)$$

where $a>0$, $b,c,d\in\mathbb{R}$, $x\in\Omega$, $t>0$ (a non-linear ex: $\partial_t u = \partial_x(u\partial_x u)+u^2$)

- Initial Conditions

$u(x,0)=u_0(x)$ $x\in\Omega$ (tells us what solution looks like at $t=0$)

- Boundary Conditions

tells us what solution looks like at boundaries of Ω

$\Omega=(0,1)$

at $x=0$: $\alpha_0(t)u+\alpha_1(t)\partial_x u=\alpha_2(t)$ $\qquad \alpha_0\geq 0,\ \alpha_1\leq 0,\ \alpha_0-\alpha_1>0$

at $x=1$: $\beta_0(t)u+\beta_1(t)\partial_x u=\beta_2(t)$ $\qquad \beta_0\geq 0,\ \beta_1\geq 0,\ \beta_0+\beta_1>0$

The Heat equation ($a(x,t)=1,\ b=c=d=0$)

$$* \quad \begin{cases} \partial_t u = \partial_{xx}u, & t>0,\ 0<x<1 \\ u(0,t)=u(1,t)=0, & t\geq 0 \\ u(x,0)=u_0(x), & 0\leq x\leq 1 \end{cases}$$

Solving the Heat equation

Separation of variables. Assume the solution has the form

$$u(x,t)=f(x)g(t)$$

Plug into $\partial_t u=\partial_{xx}u$:

$$f(x)g'(t)=f''(x)g(t)$$

$$\iff \frac{g'(t)}{g(t)}=\frac{f''(x)}{f(x)}=\text{constant}=-k^2$$

So

$$g'(t) + k^2 g(t) = 0 \iff g(t) = C_1 \exp(-k^2 t)$$
$$f''(x) + k^2 f(x) = 0 \iff f(x) = C_2 \cos(kx) + C_2 \sin(kx)$$

So

$$u(x,t) = (C_2 \cos(kt) + C_3 \sin(kt)) \exp(-k^2 t).$$

Use boundary conditions to find constants

$$0 = u(0,t) = C_2 \exp(-k^2 t) \quad \Rightarrow C_2 = 0$$
$$0 = u(1,t) = C_3 \sin(k) \exp(-k^2 t) \Rightarrow \cancel{C_3 = 0}$$

take $k = m\pi$, $m \in \mathbb{N} \setminus \{0\}$

Then

$$u_m(x,t) = a_m \sin(m\pi x) \exp(-(m\pi)^2 t).$$

u_m satisfies the PDE and BCs, so linear combinations do also.

$$u(x,t) = \sum_{m=1}^{\infty} a_m \exp(-(m\pi)^2 t) \sin(\pi m x).$$

Use the initial conditions to find a_m. At $t=0$, $u(x,0) = u_0(x)$:

$$\sum_{m=1}^{\infty} a_m \sin(m\pi x) = u_0(x).$$

a_m are the coefficients of the Fourier sine series exponsion of $u_0(x)$.

Fourier sine series of a function $f(x)$ on $x \in [-\ell, \ell]$:

$$f(x) = \frac{1}{2} a_0 + \sum_{m=1}^{\infty} \left(b_m \cos\left(\frac{m\pi x}{\ell}\right) + a_m \sin\left(\frac{m\pi x}{\ell}\right) \right)$$

where

$$b_m = \frac{1}{\ell} \int_{-\ell}^{\ell} f(x) \cos\left(\frac{m\pi x}{\ell}\right) dx, \quad a_m = \frac{1}{\ell} \int_{-\ell}^{\ell} f(x) \sin\left(\frac{m\pi x}{\ell}\right) dx$$

$x \in [0,1]$. Our solution is 0 at boundaries, so we can extend our function as an anti-symmetric function

$$u_0(x) = -u_0(-x) \text{ on } x \in [-1, 0].$$

We get $b_m = 0$ and

$$a_m = \frac{1}{\ell} \int_{-\ell}^{\ell} u_0(x) \sin(m\pi x) dx$$
$$= -\frac{1}{\ell} \int_{\ell}^{0} u_0(-z) \sin(m\pi(-z)) dz + \frac{1}{\ell} \int_{-\ell}^{\ell} u_0(x) \sin(m\pi x) dx$$
$$= 2 \int_0^1 u_0(x) \sin(m\pi x) dx$$

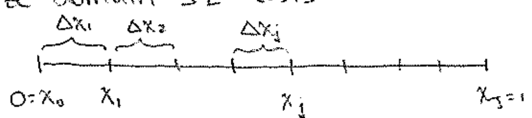
Hence

$$u(x,t) = \sum_{m=1}^{\infty} a_m \exp(-(\pi m)^2 t)\sin(m\pi x).$$

2016 02 23

Finite Difference Methods

Step 1: Discretize domain $\Omega = [0,1]$



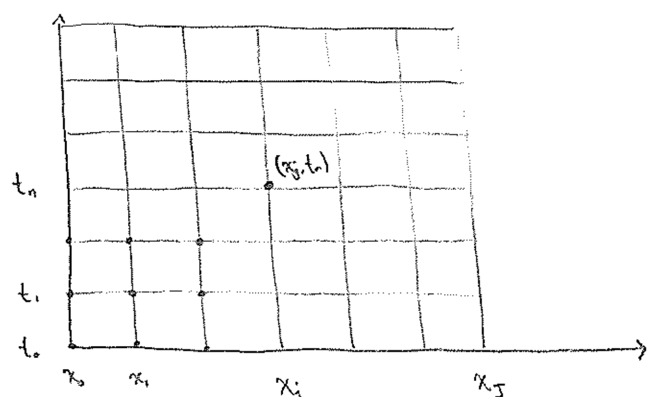
Set $\Delta x_j = x_j - x_{j-1}$. If $\Delta x_j = \Delta x$ $\forall j$ then we have a uniform grid. As the grid has $J+1$ points this means $\Delta x = \frac{1}{J}$.

Step 2: Discretize time interval $t \in [0, T_F]$

$$0 = t_0 < t_1 < t_2 < \cdots$$

$$\Delta t_n = t_n - t_{n-1}$$

Assuming a uniform timestep: $\Delta t_n = \Delta t$, $t_n = n\Delta t$



In finite differences, we seek approximations to $\partial_t u = \partial_{xx} u$ in all mesh points: $U_j^n \approx u(x_j, t_n)$.

Notation:

- U is the numerical solution
- U_j^n is the numerical solution at $t = t_n$ and $x = x_j$
- u is the exact solution
- u_j^n or $u(x_j, t_n)$ is the exact solution at $x = x_j, t = t_n$

Step 3: Apply finite differences

Main idea: replace derivatives by linear combinations of discrete function values

Example:

$$\frac{\partial u}{\partial t}(x,t) = \lim_{h\downarrow 0} \frac{u(x,t+h)-u(x,t)}{h}$$

This motivates the approximation

$$\frac{\partial u}{\partial t}(x_j,t_n) \approx \frac{u(x_j,t_n+\Delta t)-u(x_j,t_n)}{\Delta t}$$

Representation is not unique

$$\frac{\partial u}{\partial t} u(x,t) = \lim_{h\downarrow 0} \frac{u(x,t)-u(x,t-h)}{h}$$

$$= \lim_{h\downarrow 0} \frac{u(x,t+\frac{1}{2}h)-u(x,t-\frac{1}{2}h)}{h}$$

$$= \lim_{h\downarrow 0} \frac{u(x,t+h)-u(x,t-h)}{2h}$$

A summary of finite differences:

- forward differences

$$\frac{\partial u}{\partial t}(x_j,t_n) \approx \frac{u(x_j,t_n+\Delta t)-u(x_j,t_n)}{\Delta t}$$

$$\frac{\partial u}{\partial x}(x_j,t_n) \approx \frac{u(x_j+\Delta x,t_n)-u(x_j,t_n)}{\Delta x}$$

- backward differences

$$\frac{\partial u}{\partial t}(x_j,t_n) \approx \frac{u(x_j,t_n)-u(x_j,t_n-\Delta t)}{\Delta t}$$

$$\frac{\partial u}{\partial x}(x_j,t_n) \approx \frac{u(x_j,t_n)-u(x_j-\Delta x,t_n)}{\Delta x}$$

- central differences 1

$$\frac{\partial u}{\partial t}(x_j,t_n) \approx \frac{u(x_j,t_n+\Delta t)-u(x_j,t_n-\Delta t)}{2\Delta t}$$

$$\frac{\partial u}{\partial x}(x_j,t_n) \approx \frac{u(x_j+\Delta x,t_n)-u(x_j-\Delta x,t_n)}{2\Delta x}$$

- central differences 2

$$\frac{\partial u}{\partial t}(x_j,t_n) \approx \frac{u(x_j,t_n+\frac{1}{2}\Delta t)-u(x_j,t_n-\frac{1}{2}\Delta t)}{\Delta t}$$

$$\frac{\partial u}{\partial x}(x_j,t_n) \approx \frac{u(x_j+\frac{1}{2}\Delta x,t_n)-u(x_j-\frac{1}{2}\Delta x,t_n)}{\Delta x}$$

What about
$$\frac{\partial^2 u}{\partial x^2}(x_j, t_n)?$$
We have
$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) \approx \frac{\partial}{\partial x}\left(\frac{u(x_j + \frac{1}{2}\Delta x, t_n) - u(x_j - \frac{1}{2}\Delta x, t_n)}{\Delta x}\right)$$
$$= \frac{1}{\Delta x}\frac{\partial u}{\partial x}(x_j + \tfrac{1}{2}\Delta x, t_n) - \frac{1}{\Delta x}\frac{\partial u}{\partial x}(x_j - \tfrac{1}{2}\Delta x, t_n)$$
$$\approx \frac{1}{\Delta x}\left(\frac{u(x_j + \Delta x, t_n) - u(x_j, t_n)}{\Delta x}\right) - \frac{1}{\Delta x}\left(\frac{u(x_j, t_n) - u(x_j - \Delta x, t_n)}{\Delta x}\right)$$
$$= \frac{u(x_j + \Delta x, t_n) - 2u(x_j, t_n) + u(x_j - \Delta x, t_n)}{(\Delta x)^2}$$

Back to the heat equation $\partial_t u = \partial_{xx} u$. We use
$$\partial_t u(x_j, t_n) \approx \frac{u(x_j, t_n + \Delta t) - u(x_j, t_n)}{\Delta t}$$
$$\partial_x u(x_j, t_n) \approx \frac{u(x_j + \Delta x, t_n) - 2u(x_j, t_n) + u(x_j - \Delta x, t_n)}{(\Delta x)^2}$$
Hence
$$\frac{u(x_j, t_n - \Delta t) - u(x_j, t_n)}{\Delta t} \approx \frac{u(x_j + \Delta x, t_n) - 2u(x_j, t_n) + u(x_j - \Delta x, t_n)}{(\Delta x)^2}$$
This is for the exact solution u. So for the approximate solution we have
$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{(\Delta x)^2},$$
which is valid for $j = 1, 2, \dots, J-1$. At boundary points, we use the boundary conditions $u(0,t) = u(1,t) = 0$. Hence $U_0^n = U_J^n = 0$. Finally, for the initial condition, set $U_j^0 = u_0(x_j)$ for all j.

2016 02 25

Let $\nu = \Delta t / (\Delta x)^2$. Then we have
$$U_j^{n+1} = U_j^n + \frac{\Delta t}{(\Delta x)^2}\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right) = U_j^n + \nu\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right).$$
This is an <u>explicit</u> finite difference scheme.

Back to the heat equation $\partial_t u = \partial_{xx} u$. Use central differences to approximate $\partial_{xx} u$:

$$\partial_{xx} u(x_j, t_n) \approx \frac{u(x_j+\Delta x, t) - 2u(x_j, t) + u(x_j-\Delta x, t)}{(\Delta x)^2}$$

Approximate the PDE by

$$\frac{\partial U_j(t)}{\partial t} = \frac{U_{j+1}(t) - 2U_j(t) + U_{j-1}(t)}{(\Delta x)^2} = f(U_{j-1}(t), U_j(t), U_{j+1}(t)). \qquad (*)$$

Apply Euler's method to this to find

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = f(U_{j-1}, U_j, U_{j+1}) = \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{(\Delta x)^2}.$$

Example Consider the 2-stage ERK method

$$\begin{array}{c|cc} 0 & & \\ \frac{2}{3} & \frac{2}{3} & \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}$$

applied to $(*)$:

* $(k_1)_j = U_j^n$
* $(k_2)_j = U_j^n + \frac{2}{3}\left(\frac{\Delta t}{\Delta x^2}\right)\left((k_1)_{j+1} - 2(k_1)_j + (k_1)_{j-1}\right)$
* $U_j^{n+1} = U_j^n + \frac{1}{4}\frac{\Delta t}{\Delta x^2}\left((k_1)_{j+1} - 2(k_1)_j + (k_1)_{j-1}\right) + \frac{3}{4}\frac{\Delta t}{\Delta x^2}\left((k_2)_{j+1} - 2(k_2)_j + (k_2)_{j-1}\right).$

Truncation error

Let u be the solution to our model problem.
If we use central differences to approximate $\partial_{xx} u$ and forward differences for $\partial_t u$ then the finite difference method on interior points is given by

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{\Delta x^2}.$$

The truncation error is the error that we make if we replace U by u in the finite difference method:

$$T(x,t) = \frac{u(x, t+\Delta t) - u(x,t)}{\Delta t} - \frac{u(x+\Delta x, t) - 2u(x,t) + u(x-\Delta x, t)}{\Delta x^2}$$

Can we say anything about the size of $T(x,t)$?
If T is huge, then U is probably not a good approximation

Use Taylor series expansion:

$$* \; u(x,t+\Delta t) = u(x,t) + \Delta t\, u_t + \tfrac{1}{2}(\Delta t)^2 u_{tt} + \tfrac{1}{6}(\Delta t)^3 u_{ttt} + \cdots$$

$$* \; u(x+\Delta x,t) = u(x,t) + \Delta x\, u_x + \tfrac{1}{2}(\Delta x)^2 u_{xx} + \tfrac{1}{6}(\Delta x)^3 u_{xxx} + \cdots$$

$$* \; u(x-\Delta x,t) = u(x,t) - \Delta x\, u_x + \tfrac{1}{2}(\Delta x)^2 u_{xx} - \tfrac{1}{6}(\Delta x)^3 u_{xxx} + \cdots$$

Substituting into the expression for $T(x,t)$:

$$T(x,t) = \left(u_t + \tfrac{1}{2}\Delta t\, u_{tt} + \cdots\right) - \left(u_{xx} + \frac{\Delta x^2}{12} u_{xxxx} + \cdots\right)$$

$$= \underbrace{u_t - u_{xx}}_{=0} + \tfrac{1}{2}\Delta t\, u_{tt} - \tfrac{1}{12}\Delta x^2 u_{xxxx} + \cdots$$

$$= \tfrac{1}{2}\Delta t\, u_{tt} - \tfrac{1}{12}\Delta x^2 u_{xxxx} + \text{h.o.t.}$$

Let $\eta \in (t, t+\Delta t)$, $\xi \in (x-\Delta x, x+\Delta x)$ be such that

$$T(x,t) = \tfrac{1}{2}\Delta t\, u_{tt}(x,\eta) - \tfrac{1}{12}(\Delta x)^2 u_{xxxx}(\xi,t).$$

We can now bound the truncation error:

$$|T(x,t)| = \left|\tfrac{1}{2}\Delta t\, u_{tt}(x,\eta) - \tfrac{1}{12}(\Delta x)^2 u_{xxxx}(\xi,t)\right|$$

$$\leq \tfrac{1}{2}\Delta t\, |u_{tt}(x,\eta)| + \tfrac{1}{12}\Delta x^2 |u_{xxxx}(\xi,t)|.$$

Say $x \in \Omega$, $t \in [0,T_F]$. Let

$$M_{tt} = \max_{\substack{x\in\Omega \\ t\in[0,T_F]}} |u_{tt}| \qquad M_{xxxx} = \max_{\substack{x\in\Omega \\ t\in[0,T_F]}} .$$

Then

$$|T(x,t)| \leq \tfrac{1}{2}\Delta t\, M_{tt} + \tfrac{1}{12}\Delta x^2 M_{xxxx} \quad \text{———} (*)$$

$$= \tfrac{1}{2}\Delta t\left(M_{tt} + \frac{1}{6\nu} M_{xxxx}\right).$$

First order accuracy in time if ν is fixed.

<u>Increasing accuracy</u>

From $u_t = u_{xx}$ we have

$$\left.\begin{array}{ll}
\cdot\; \partial_x(u_t) = \partial_x(u_{xx}) \Rightarrow & u_{xt} = u_{xxx} \\
\cdot\; \partial_t(u_t) = \partial_x(u_{xx}) & u_{tt} = u_{txx} \\
\cdot\; \partial_{xx}(u_t) = \partial_{xx}(u_{xx}) & u_{xxt} = u_{xxxx} \\
?? \longrightarrow \cdot\; \partial_{xt}(u_t) = \partial_{xt}(u_{xx}) & u_{txx} = u_{xxt}
\end{array}\right\} \quad u_{tt} = u_{xxxx}$$

So

$$T(x,t) = \left(\tfrac{1}{2}\Delta t - \tfrac{1}{12}\Delta x^2\right) u_{xxxx}(\xi,\eta) = \tfrac{1}{2}\Delta t\left(1 - \tfrac{1}{6\nu}\right)u_{xxxx}(x,t) + O(\Delta t^2).$$

If $\nu = 1/6$ then this is $O(\Delta t^2)$

2016 02 29

Looking back at (*), we have:

- $T(x,t) \to 0$ as $\Delta t \to 0$, $\Delta x \to 0$ independent from each other $\Rightarrow$ scheme is unconditionally consistent.
- If we fix ν, then the truncation error is order $O(\Delta t)$ $\Rightarrow$ scheme is first order accurate.

Convergence: Fix ν. Define the mesh such that $\Delta t \to 0$ and $\Delta x \to 0$. The scheme is convergent if for any fixed point (x^*, t^*) in domain $(0,1)\times(0,T_F)$ we have

$$\left.\begin{array}{l} x_j \to x^* \\ t^n \to t^* \end{array}\right\} \Rightarrow U_j^n \to u(x^*,t^*)$$

Convergence of FD method ($\nu \le \frac{1}{2}$)

Define the error as $e_j^n = U_j^n - u(x_j, t_n)$, where u satisfies $\partial_t u = \partial_{xx} u$ and U_j^n satisfies

$$U_j^{n+1} = U_j^n + \nu\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right). \qquad \text{——— (a)}$$

Substitute u into the FD equation:

$$u_j^{n+1} = u_j^n + \nu\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right) + \Delta t\, T(x,t) \qquad \text{——— (b)}$$

Subtract (b) from (a):

$$U_j^{n+1} - u_j^{n+1} = U_j^n - u_j^n + \nu\left(U_{j+1}^n - u_{j+1}^n - 2U_j^n + 2u_j^n + U_{j-1}^n - u_{j-1}^n\right) - \Delta t\, T(x_j, t_n)$$

$$\Rightarrow \quad e_j^{n+1} = e_j^n + \nu\left(e_{j+1}^n - 2e_j^n + e_{j-1}^n\right) - \Delta t\, T(x_j, t_n)$$

$$= (1-2\nu)e_j^n + \nu e_{j+1}^n + \nu e_{j-1}^n - \Delta t\, T(x_j, t_n)$$

$$\Rightarrow \quad |e_j^{n+1}| = \left|(1-2\nu)e_j^n + \nu e_{j+1}^n + \nu e_{j-1}^n - \Delta t\, T(x_j, t_n)\right|$$

$$\le (1-2\nu)|e_j^n| + \nu|e_{j+1}^n| + \nu|e_{j-1}^n| + \Delta t\,|T(x_j,t_n)| \qquad \text{(since } \nu \le \tfrac{1}{2}\text{)}$$

$$\le (1-2\nu)\max_j |e_j^n| + \nu \max_j |e_j^n| + \nu \max_j |e_j^n| + \Delta t \max_j |T_j^n|$$

Define

$$E^n = \max\{|e_j^n|\,;\ j = 0, \ldots, J\}, \quad \bar{T} = \max_{j,n} |T_j^n|.$$

Then

$$E^{n+1} \le (1-2\nu)E^n + \nu E^n + \nu E^n + \Delta t \bar{T} = E^n + \bar{T}\Delta t.$$

At $t=0$ the initial values for U_j^n are defined by $u_0(x)$, so $E^0 = 0$. Now:

$$\left.\begin{aligned} t_0 &: E^0 = 0 \\ t_1 &: E^1 \le E^0 + \bar{T}\Delta t = \bar{T}\Delta t \\ t_2 &: E^2 \le E^1 + \bar{T}\Delta t = 2\bar{T}\Delta t \\ t_3 &: E^3 \le E^2 + \bar{T}\Delta t = 3\bar{T}\Delta t \\ &\vdots \end{aligned}\right\} \Longrightarrow E^n \le n\bar{T}\Delta t \quad \text{(induction)}$$

Recall

$$|T(x,t)| \le \tfrac{1}{2}\Delta t \left(M_{tt} + \tfrac{1}{6\nu} M_{xxxx}\right).$$

Hence

$$\begin{aligned} E^n &\le \tfrac{1}{2}\Delta t \left(M_{tt} + \tfrac{1}{6\nu} M_{xxxx}\right) n\Delta t \\ &\le \tfrac{1}{2}\Delta t \left(M_{tt} + \tfrac{1}{6\nu} M_{xxxx}\right) T_F \end{aligned}$$

So $E^n \to 0$ as $\Delta t \to 0$ and so we have convergence.

Now we look at convergence in slightly more general terms.

Def) Refinement path: a sequence of pairs of mesh sizes Δx and (Δt), each of which tends to zero: $\{((\Delta x)_i, (\Delta t)_i),\ i = 0,1,2,\ldots,\ (\Delta x)_i, (\Delta t)_i \to 0\}$.

Examples are

$$\nu_i = \frac{(\Delta t)_i}{(\Delta x)_i^2}, \qquad \nu_i = \frac{(\Delta t)_i}{(\Delta x)_i}.$$

Theorem 2.1: If a refinement path satisfies $\nu_i = \frac{1}{2}$ for all sufficiently large values of i, and positive numbers n_i, j_i are such that

$$n_i(\Delta t)_i \to t > 0, \quad j_i(\Delta x)_i \to x$$

and if $|u_{xxxx}| \le M_{xxxx}$ then the FD solution U_j^n will converge to u.

Midterm

proofs of conv

sec. 3.8

under

2016 03 01

We've seen

- consistent (truncation error $T \to 0$ as $\Delta x \to 0, \Delta t \to 0$) (§2.5)
- convergent ($\nu \le \frac{1}{2}$) (§2.6)

We'll see now

- stability (§2.7)

Suppose $U_j^n = \lambda^n \exp(ijk\Delta x)$. Then

$$U_j^{n+1} = \lambda^{n+1}\exp(ijk\Delta x) = \lambda U_j^n$$
$$U_{j+1}^n = \lambda^n \exp(ik(j+1)\Delta x) = U_j^n \exp(ik\Delta x)$$
$$U_{j-1}^n = \lambda^n \exp(ik(j-1)\Delta x) = U_j^n \exp(-ik\Delta x).$$

Substitute these into $U_j^{n+1} = U_j^n + \nu(U_{j+1}^n - 2U_j^n + U_{j-1}^n)$ to get

$$\lambda U_j^n = U_j^n + \nu(U_j^n \exp(ik\Delta x) - 2U_j^n + U_j^n \exp(-ik\Delta x))$$
$$\Rightarrow \lambda(k) = 1 + \nu(\exp(ik\Delta x) - 2 + \exp(-ik\Delta x))$$
$$= 1 + \nu(\cos(k\Delta x) + \cancel{i\sin(k\Delta x)} - 2 + \cos(k\Delta x) - \cancel{i\sin(k\Delta x)})$$
$$= 1 + 2\nu(\cos(k\Delta x) - 1) \qquad (\sin^2(x) = \tfrac{1}{2}(1-\cos(2x)))$$
$$= 1 - 4\nu\sin^2(\tfrac{1}{2}k\Delta x)$$

This $\lambda(k)$ is called the amplification factor for the mode with index k.
So $U_j^n = \lambda^n \exp(ijk\Delta x)$ is a solution of the FD scheme provided that

$$\lambda(k) = 1 - 4\nu\sin^2(\tfrac{1}{2}k\Delta x).$$

Then

$$U_j^n = \sum_{k=-\infty}^{\infty} A_k \exp(ikj\Delta x)(\lambda(k))^n.$$

Compare this to the exact solution

$$u(x_j, t_n) = \sum_{k=-\infty}^{\infty} A_k \exp(ikj\Delta x)(\exp(-k^2\Delta t))^n.$$

For low frequency modes (k small) the numerical expansion is a good approximation to the exact solution:

- $\exp(-k^2t) = (1 - k^2\Delta t) + \frac{1}{2}k^4(\Delta t)^2 + \cdots$

- $\lambda(k) = 1 - 2\nu(1-\cos(k\Delta x)) = 1 - 2\nu(1 - 1 + \frac{1}{2}(k\Delta x)^2 - \frac{1}{24}(k\Delta x)^4 + \cdots)$

$$= (1 - k^2\Delta t) + \tfrac{1}{12}k^4\Delta t(\Delta x)^2 + \cdots$$

What about large k?

- For u the modes rapidly decay
- For U, if $|\lambda(k)| > 1$ then the Fourier modes will grow unboundedly.

So we require $|\lambda(k)| \le 1 \ \forall k$ for stability.

$$\lambda = 1 - 4\nu\sin^2(\tfrac{1}{2}k\Delta x)$$
$$|\lambda| = |1 - 4\nu\sin^2(\tfrac{1}{2}k\Delta x)| \le 1$$
$$\Leftrightarrow 4\nu\sin^2(\tfrac{1}{2}k\Delta x) \le 2 \Leftrightarrow 4\nu \le 2 \Leftrightarrow \nu \le \tfrac{1}{2}.$$

Von Neumann stability condition

For stability we require $|\lambda(k)| \le 1 + K\Delta t$ (we focus on $|\lambda(k)| \le 1$).

Extension of convergence using Fourier Analysis

Previous convergence proof required us to assume

- the solution is smooth such that $\partial_{tt}u$ and $\partial_{xxxx}u$ are bounded.

Now we can relax this assumption and assume only that

- u is continuous
- the initial condition $u_0(x)$ has absolutely convergent Fourier series

Proof of convergence 2: Suppose $\nu \le \frac{1}{2}$. Fix ν. The error is given by

$$e_j^n = U_j^n - u_j^n = \sum_{m=-\infty}^{\infty} A_m \exp(ijm\pi\Delta x)(\lambda(m\pi))^n - \sum_{m=-\infty}^{\infty} A_m \exp(ijm\pi\Delta x)\exp(-m^2\pi^2 n\Delta t)$$

$$= \sum_{m=-\infty}^{\infty} A_m \exp(ijm\pi\Delta x)\Big((\lambda(m\pi))^n - \exp(-m^2\pi^2 n\Delta t)\Big).$$

Split the infinite sum into two parts. Given $\varepsilon > 0$, choose m_0 such that

$$\sum_{|m|>|m_0|} |A_m| \le \frac{1}{4}\varepsilon.$$

Note that if $|\lambda_1| \le 1$ and $|\lambda_2| \le 1$, then $|\lambda_1^n - \lambda_2^n| \le n|\lambda_1 - \lambda_2|$.

Now

$$|e_j^n| = |U_j^n - u_j^n|$$

$$\le \left|\sum_{|m|\le m_0} A_m \exp(im\pi j\Delta x)\Big((\lambda(m\pi))^n - \exp(-m^2\pi^2 n\Delta t)\Big)\right| + \frac{1}{2}\varepsilon$$

$$\le \sum_{|m|\le m_0} |A_m| \left|(\lambda(m\pi))^n - (\exp(-m^2\pi^2\Delta t))^n\right| + \frac{1}{2}\varepsilon$$

$$\le \sum_{|m|\le m_0} n|A_m| \underbrace{|\lambda(m\pi) - \exp(-m^2\pi^2\Delta t)|}_{\le C(\nu)(m^2\pi^2\Delta t)^2 \text{ (Taylor series)}} + \frac{1}{2}\varepsilon$$

$$\le \sum_{|m|\le m_0} |A_m| n C(\nu)(m^2\pi^2\Delta t)^2 + \frac{1}{2}\varepsilon$$

$$\le T_F \pi^4 C(\nu)\Delta t \underbrace{\sum_{|m|\le m_0} |A_m| m^4}_{\le M} + \frac{1}{2}\varepsilon.$$

2016 03 03

So letting $K = T_F\pi^4 C(\nu) M$ we have

$$|e_j^n| \le K\Delta t + \tfrac{1}{2}\varepsilon.$$

As $\Delta t \to 0$, $|e_j^n| \le \tfrac{1}{2}\varepsilon \quad \forall (x_j, t_n) \in [0,1] \times [0, T_F]$.

Advantages of this analysis: we do not have to assume that u is four times differentiable.

Recall that we are solving the heat equation via
- forward differences in time
- central differences in space

We have
- consistency
- convergence
- stability

provided $\nu \le \frac{1}{2}$. But $\Delta t \le \frac{1}{2}(\Delta x)^2$ is a severe restriction!

Still using central differences in space,

$$\frac{\partial U_j(t)}{\partial t} = \frac{U_{j+1}(t) - 2U_j(t) + U_{j-1}(t)}{(\Delta x)^2}.$$

- Euler's method gives $\nu \le \frac{1}{2}$
- Backward Euler method:

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}}{(\Delta x)^2} \qquad (j = 1, 2, \dots, J-1)$$

$$U_0^{n+1} = U_J^{n+1} = 0$$

$$U_j^0 = u_0(x_j)$$

So

$$U_j^{n+1} = U_j^n + \nu\left(U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}\right)$$

$$\Longleftrightarrow \quad \underbrace{-\nu U_{j-1}^{n+1} + (1+2\nu)U_j^{n+1} - \nu U_{j+1}^{n+1}}_{\text{unknowns}} = U_j^n$$

To solve this equation we will write in matrix form:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & & & 0 \\ -\nu & 1+2\nu & -\nu & 0 & & & & 0 \\ 0 & \ddots & \ddots & \ddots & & & & \vdots \\ \vdots & & & & & & & \\ 0 & \cdots & 0 & -\nu & 1+2\nu & -\nu & 0 \cdots & 0 \\ \vdots & & & \ddots & \ddots & \ddots & & 0 \\ & & & & 0 & -\nu & 1+2\nu & -\nu \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} U_0^{n+1} \\ U_1^{n+1} \\ \vdots \\ \\ U_j^{n+1} \\ \vdots \\ U_{J-1}^{n+1} \\ U_J^{n+1} \end{bmatrix} = \begin{bmatrix} 0 \\ U_1^n \\ \vdots \\ \\ U_j^n \\ \vdots \\ U_{J-1}^n \\ 0 \end{bmatrix}$$

"how do we solve this?" - prof
"invert A" - student 1
"Sure." - prof

→ "What if you don't have MATLAB?" - prof
"Get it." - student 1
"..." - prof
"Guess and check your answer?" - student 2

So we have $AU^{n+1} = B$ where A is a sparse matrix:

$$A(1,1) = 1$$
$$\left.\begin{aligned} A(j,j-1) &= -v \\ A(j,j) &= 1+2v \\ A(j,j+1) &= -v \end{aligned}\right\} j = 1,2,\dots,J-1$$
$$A(J,J) = 1$$
$$A(i,j) = 0 \quad \text{otherwise}$$

To solve
- $U^{n+1} = A \backslash B$ in Matlab
- Thomas algorithm
- Iterative methods

Thomas Algorithm

Assume a system of the form
- $-a_j U_{j-1} + b_j U_j - c_j U_{j+1} = d_j \quad j = 1,2,\dots,J-1$
- $U_0 = U_J = 0$
- $a_j, b_j, c_j > 0$ and $b_j > a_j + c_j$ (for diagonal dominance)

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & & 0 \\ -a_1 & b_1 & -c_1 & & & & \\ 0 & \ddots & \ddots & \ddots & & & \\ \vdots & & -a_j & b_j & c_j & & \\ & & & \ddots & \ddots & \ddots & 0 \\ & & & & -a_{J-1} & b_{J-1} & c_{J-1} \\ 0 & & \cdots & 0 & 0 & 0 & \end{bmatrix} \begin{bmatrix} U_0 \\ U_1 \\ \vdots \\ U_j \\ \vdots \\ U_{J-1} \\ U_J \end{bmatrix} = \begin{bmatrix} 0 \\ d_1 \\ \vdots \\ d_j \\ \vdots \\ d_{J-1} \\ 0 \end{bmatrix}$$

We have $U_0 = U_J = 0$ so we can eliminate the outter rows/columns. We get the augmented matrix

$$\left[\begin{array}{cccc|c} b_1 & -c_1 & & & d_1 \\ -a_2 & b_2 & -c_2 & & d_2 \\ & \ddots & \ddots & \ddots & \vdots \\ & & -a_{J-1} & b_{J-1} & d_{J-1} \end{array}\right] \begin{array}{l} * \frac{1}{b_1} \\ +a_2 * \text{row } 1 \\ \\ \\ \end{array}$$

$$\left[\begin{array}{ccccc|c} 1 & -c_1/b_1 & & & & d_1/b_1 \\ 0 & (b_2 - a_2 c_1/b_1) & -c_2 & & & d_2 + a_2 d_1/b_1 \\ 0 & -a_3 & b_3 & -c_3 & & d_3 \\ & & \ddots & \ddots & \ddots & \vdots \end{array}\right]$$

Let $e_1 = c_1/b_1$, $f_1 = d_1/b_1$.

$$\left[\begin{array}{cccc|c} 1 & -e_1 & & & f_1 \\ & 1 & \frac{-c_2}{b_2-a_2e_1} & & \frac{d_2+a_2f_1}{b_2-a_2e_1} \\ & -a_3 & b_3 & -c_3 \ddots & \\ & & \ddots & \ddots & \end{array}\right]$$

Let $e_2 = c_2/(b_2-a_2e_1)$, $f_2 = (d_2+a_2f_1)/(b_2-a_2e_1)$. After k iterations

$$\left[\begin{array}{cccccc|c} 1 & -e_1 & & & & & f_1 \\ & 1 & -e_2 & & & & f_2 \\ & & 1 & -e_3 & & & \vdots \\ & & & \ddots \; 1 & -e_k & & f_k \\ & & & -a_{k+1} & b_{k+1} & -c_{k+1} & d_{k+1} \\ & & & & \ddots & \ddots & \vdots \\ & & & & -a_{J-1} & b_{J-1} & d_{J-1} \end{array}\right]$$

Looking at rows k and $k+1$:

$$U_k - e_k U_{k+1} = f_k, \quad -a_{k+1}U_k + b_{k+1}U_{k+1} - c_{k+1}U_{k+2} = d_{k+1}$$

$$\Rightarrow U_{k+1} - \underbrace{\frac{c_{k+1}}{b_{k+1}-a_{k+1}e_k}}_{e_{k+1}} U_{k+2} = \underbrace{\frac{d_{k+1}+a_{k+1}f_k}{b_{k+1}-a_{k+1}e_k}}_{f_{k+1}}$$

So, let $e_0=0$, $f_0=0$. Compute

$$e_j = \frac{c_j}{b_j-a_je_{j-1}}, \quad f_j = \frac{d_j+a_jf_{j-1}}{b_j-a_je_{j-1}}, \quad j=1,\dots,J-1$$

Obtain U_k from $U_k - e_kU_{k+1} = f_k$ starting at $k=J-1, J-2, \dots, 1$

So we know 2016 02 07

$$U_J = 0$$
$$U_{J-1} = f_{J-1} + e_{J-1}U_J$$
$$U_{J-2} = f_{J-2} + e_{J-2}U_{J-1}$$
$$\vdots$$

For backward Euler + central differences FD method for the heat equation:

$$-\nu U_{j-1}^{n+1} + (1+2\nu)U_j^{n+1} - \nu U_{j+1}^{n+1} = U_j^n \qquad j=1,\dots,J-1$$

$$a_j=\nu, \quad b_j = 1+2\nu, \quad c_j=\nu, \quad d_j = U_j^n$$

$$a_j, b_j, c_j > 0 \checkmark \qquad b_j > a_j + c_j \checkmark$$

We can apply the Thomas algorithm. $e_0=0$, $f_0=0$. Compute

$$e_j = \frac{\nu}{(1+2\nu)-\nu e_{j-1}}, \quad f_j = \frac{U_j^n + \nu f_{j-1}}{(1+2\nu)-\nu e_{j-1}}, \quad j=1,\dots,J-1$$

Given $U_J^{n+1}=0$,

```
for  k = J-1 : 1
     U_k^{n+1} = f_k + e_k U_{k+1}^{n+1}
end.
```

For foward Euler, we required $\Delta t \le \frac{1}{2}(\Delta x)^2$ for stability. What about backward Euler?

Consider Fourier modes

$$U_j^n = \lambda^n \exp(ikj\Delta x).$$

Then

$$U_{j-1}^{n+1} = \lambda U_j^n \exp(-ik\Delta x),$$
$$U_{j+1}^{n+1} = \lambda U_j^n \exp(ik\Delta x),$$
$$U_j^{n+1} = \lambda U_j^n.$$

Substitute into the FD scheme:

$$-\nu \lambda U_j^n \exp(-ik\Delta x) + (1+2\nu)\lambda U_j^n - \nu \lambda U_j^n \exp(ik\Delta x) = U_j^n$$
$$\Rightarrow \lambda\left(-\nu \exp(-ik\Delta x) + (1+2\nu) - \nu \exp(ik\Delta x)\right) = 1$$
$$\Rightarrow \lambda - 1 = \lambda\nu\left(\exp(-ik\Delta x) - 2 + \exp(ik\Delta x)\right)$$
$$\Rightarrow \lambda - 1 = 2\lambda\nu\left(\cos(k\Delta x) - 1\right)$$
$$\Rightarrow \lambda - 1 = -4\nu\lambda \sin^2\left(\tfrac{1}{2}k\Delta x\right)$$

So the amplification factor is

$$\lambda(k) = \frac{1}{1 + 4\nu \sin^2(\frac{1}{2}k\Delta x)}.$$

For stability, we need $|\lambda(k)| < 1$. Well,

$$\left|\frac{1}{1+4\nu\sin^2(\frac{1}{2}k\Delta x)}\right| < 1 \quad \forall \nu > 0.$$

Therefore the backward Euler method is <u>unconditionally</u> stable! We can take Δt as big as we want and still have a stable scheme. <u>Caution:</u> Don't take Δt too big; we still want an accurate solution!

All our methods use

$$\frac{\partial U_j}{\partial t}(t) = \frac{U_{j+1}(t) - 2U_j(t) + U_{j-1}(t)}{(\Delta x)^2}$$

We've seen

- forward Euler
  - consistency ✓
  - convergence ✓
  - stability ($\nu \le \frac{1}{2}$) ✓
- backward Euler
  - stability ($\forall \nu$) ✓

Now we will look at the θ-method, which covers backward Euler, forward Euler, and trapezoidal rule.

$$U_j^{n+1} = U_j^n + \nu\theta\left(U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}\right) + \nu(1-\theta)\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right)$$

$$\Leftrightarrow\ -\nu\theta U_{j-1}^{n+1} + (1+2\nu\theta)U_j^{n+1} - \nu\theta U_{j+1}^{n+1} = U_j^n + (1-\theta)\nu\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right) \qquad -\ (*)$$

Observations:

- $\theta = 0 \longrightarrow$ forward Euler in time
- $\theta = 1 \longrightarrow$ backward Euler in time
- $\theta = \frac{1}{2} \longrightarrow$ Trapezoidal rule (Crank-Nicolson scheme)
- $\theta = 0$ is the only explicit method; $\theta > 0$ is implicit

2016 03 08

<u>Stability of θ-method</u>

Choose $U_j^n = \lambda^n \exp(ikj\Delta x)$ and substitute into $(*)$:

$$-\nu\theta\lambda\exp(-ik\Delta x) + (1+2\nu\theta)\lambda - \nu\theta\lambda\exp(ik\Delta x)$$
$$= 1 + (1-\theta)\nu\exp(ik\Delta x) - 2\nu(1-\theta) + (1-\theta)\nu\exp(-ik\Delta x)$$

$$\Leftrightarrow \lambda - 1 = \nu\theta\lambda\underbrace{\left(\exp(ik\Delta x) - 2 + \exp(-ik\Delta x)\right)}_{-4\sin^2(\frac{1}{2}k\Delta x)} + (1-\theta)\nu\underbrace{\left(\exp(ik\Delta x) - 2 + \exp(-ik\Delta x)\right)}_{-4\sin^2(\frac{1}{2}k\Delta x)}$$

$$= -4\nu\left(\lambda\theta + (1-\theta)\right)\sin^2\left(\tfrac{1}{2}k\Delta x\right)$$

$$\Leftrightarrow\ \lambda\left(1 + 4\nu\theta\sin^2\left(\tfrac{1}{2}k\Delta x\right)\right) = 1 - 4\nu(1-\theta)\sin^2\left(\tfrac{1}{2}k\Delta x\right)$$

The amplification factor for the θ-method is given by

$$\lambda(k) = \frac{1 - 4\nu(1-\theta)\sin^2(\frac{1}{2}k\Delta x)}{1 + 4\nu\theta\sin^2(\frac{1}{2}k\Delta x)}$$

For stability: $|\lambda(k)| < 1$. We know $\nu > 0$ and $0 \le \theta \le 1$. Hence $\lambda(k) \le 1$.
We need to check $\lambda \ge -1$.

$$\lambda(k) \ge -1$$
$$\Leftrightarrow\ 1 - 4\nu(1-\theta)\sin^2\left(\tfrac{1}{2}k\Delta x\right) \ge -\left(1 + 4\nu\theta\sin^2\left(\tfrac{1}{2}k\Delta x\right)\right)$$
$$\Leftrightarrow\ 1 - 4\nu(1-2\theta)\sin^2\left(\tfrac{1}{2}k\Delta x\right) \ge -1$$
$$\Leftrightarrow\ 4\nu(1-2\theta)\underbrace{\sin^2\left(\tfrac{1}{2}k\Delta x\right)}_{\max = 1} \le 2$$
$$\Leftrightarrow\ \nu(1-2\theta) \le \tfrac{1}{2}.$$

Some cases:

- $\theta = 0$ (forward Euler) $\longrightarrow$ stable if $\nu \le \frac{1}{2}$
- $\theta = 1$ (backward Euler) $\longrightarrow$ stable for all ν
- $\theta < \frac{1}{2}$ $\longrightarrow$ stable if $\nu \le \frac{1}{2(1-2\theta)}$ (conditionally stable)

• $\theta \geq \frac{1}{2} \longrightarrow$ stable for all ν

How accurate is the θ-method? Use Taylor series expansions. Expand around $(x_j, t_{n+1/2})$, not (x_j, t_n):

$$u_j^{n+1} = \left(u + \tfrac{1}{2}\Delta t\, u_t + \tfrac{1}{2}(\tfrac{1}{2}\Delta t)^2 u_{tt} + \tfrac{1}{6}(\tfrac{1}{2}\Delta t)^3 u_{ttt} + \cdots\right]_j^{n+1/2}$$

$$u_j^{n} = \left(u - \tfrac{1}{2}\Delta t\, u_t + \tfrac{1}{2}(\tfrac{1}{2}\Delta t)^2 u_{tt} - \tfrac{1}{6}(\tfrac{1}{2}\Delta t)^3 u_{ttt} + \cdots\right]_j^{n+1/2}$$

$$u_j^{n+1} - u_j^n = \left(\Delta t\, u_t + \tfrac{1}{24}(\Delta t)^3 u_{ttt} + \cdots\right)_j^{n+1/2}$$

$$u_{j+1}^{n+1} = \left(u + \Delta x\, u_x + \tfrac{1}{2}(\Delta x)^2 u_{xx} + \tfrac{1}{6}(\Delta x)^3 u_{xxx} + \tfrac{1}{24}(\Delta x)^4 u_{xxxx} + \tfrac{1}{120}(\Delta x)^5 u_{xxxxx} + \tfrac{1}{720}(\Delta x)^6 u_{xxxxxx} + \cdots\right)_j^{n+1}$$

$$u_{j-1}^{n+1} = \left(u - \Delta x\, u_x + \tfrac{1}{2}(\Delta x)^2 u_{xx} - \tfrac{1}{6}(\Delta x)^3 u_{xxx} + \tfrac{1}{24}(\Delta x)^4 u_{xxxx} - \tfrac{1}{120}(\Delta x)^5 u_{xxxxx} + \tfrac{1}{720}(\Delta x)^6 u_{xxxxxx} + \cdots\right)_j^{n+1}$$

$$u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1} = \left((\Delta x)^2 u_{xx} + \tfrac{1}{12}(\Delta x)^4 u_{xxxx} + \tfrac{1}{360}(\Delta x)^6 u_{xxxxxx} + \cdots\right)_j^{n+1}$$

Taylor expand now in time around $t_{n+1/2}$

$$\begin{aligned} u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1} &= \left((\Delta x)^2 u_{xx} + \tfrac{1}{12}(\Delta x)^4 u_{xxxx} + \tfrac{1}{360}(\Delta x)^6 u_{xxxxxx} + \cdots\right)_j^{n+1/2} \\ &+ \tfrac{1}{2}\Delta t\left((\Delta x)^2 u_{xxt} + \tfrac{1}{12}(\Delta x)^4 u_{xxxxt} + \cdots\right)_j^{n+1/2} \\ &+ \tfrac{1}{2}(\tfrac{1}{2}\Delta t)^2\left((\Delta x)^2 u_{xxtt} + \cdots\right)_j^{n+1/2} \end{aligned}$$

Similar work would show

$$\begin{aligned} u_{j+1}^{n} - 2u_j^{n} + u_{j-1}^{n} &= \left((\Delta x)^2 u_{xx} + \tfrac{1}{12}(\Delta x)^4 u_{xxxx} + \tfrac{1}{360}(\Delta x)^6 u_{xxxxxx} + \cdots\right)_j^{n+1/2} \\ &- \tfrac{1}{2}\Delta t\left((\Delta x)^2 u_{xxt} + \tfrac{1}{12}(\Delta x)^4 u_{xxxxt} + \cdots\right)_j^{n+1/2} \\ &+ \tfrac{1}{2}(\tfrac{1}{2}\Delta t)^2\left((\Delta x)^2 u_{xxtt} + \cdots\right)_j^{n+1/2} \end{aligned}$$

Combining these yields

$$\begin{aligned} &\theta(u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}) + (1-\theta)(u_{j+1}^{n} - 2u_j^{n} + u_{j-1}^{n}) \\ &= \left((\Delta x)^2 u_{xx} + \tfrac{1}{12}(\Delta x)^4 u_{xxxx} + \tfrac{1}{360}(\Delta x)^6 u_{xxxxxx} + \cdots\right)_j^{n+1/2} \\ &+ (2\theta-1)\tfrac{1}{2}\Delta t\left((\Delta x)^2 u_{xxt} + \tfrac{1}{12}(\Delta x)^4 u_{xxxxt} + \cdots\right)_j^{n+1/2} \\ &+ \tfrac{1}{8}(\Delta t)^2(\Delta x)^2\left(u_{xxtt} + \cdots\right)_j^{n+1/2} \end{aligned}$$

The truncation error is

$$\begin{aligned} T_j^{n+1/2} &= \frac{u_j^{n+1} - u_j^n}{\Delta t} - \left(\frac{\theta}{(\Delta x)^2}(u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}) + \frac{(1-\theta)}{(\Delta x)^2}(u_{j+1}^{n} - 2u_j^{n} + u_{j-1}^{n})\right) \\ &= \underbrace{(u_t - u_{xx})}_{=0 \text{ PDE}} + \left((\tfrac{1}{2}-\theta)\Delta t\, u_{xxt} - \tfrac{1}{12}(\Delta x)^2 u_{xxxx}\right)_j^{n+1/2} \\ &+ \left(\tfrac{1}{24}(\Delta t)^2 u_{ttt} - \tfrac{1}{8}(\Delta t)^2 u_{xxtt}\right)_j^{n+1/2} \\ &+ \left(\tfrac{1}{12}(\tfrac{1}{2}-\theta)\Delta t(\Delta x)^2 u_{xxxxt} - \tfrac{1}{360}(\Delta x)^4 u_{xxxxxx}\right)_j^{n+1/2} + \text{h.o.t.} \end{aligned}$$

Observations:

- θ-method is consistent: $|T| \to 0$ as $\Delta t, \Delta x \to 0$
- if $\theta = \frac{1}{2}$ then second order accurate in time, otherwise first order

2016 03 10

Take $\theta = \frac{1}{2}$. Then

$$T_j^{n+1/2} = \left(\frac{1}{24}(\Delta t)^2 u_{ttt} - \frac{1}{12}(\Delta x)^2 u_{xxxx} - \frac{1}{8}(\Delta t)^2 u_{xxtt} + \cdots \right)_j^{n+1/2}$$

$$= -\frac{1}{12}\left((\Delta x)^2 u_{xxxx} + (\Delta t)^2 u_{ttt} \right)_j^{n+1/2} + \cdots$$

because $u_t = u_{xx}$ implies $u_{xxtt} = u_{ttt}$. So if $\theta = \frac{1}{2}$ (Crank-Nicolson) then this method is second order accurate in <u>both</u> space and time.

So far we have:

- $\theta = 0$: $O(\Delta t)$, $O(\Delta x^2)$, conditionally stable, explicit
- $\theta = 1$: $O(\Delta t)$, $O(\Delta x^2)$, unconditionally stable, implicit
- $\theta = \frac{1}{2}$: $O(\Delta t^2)$, $O(\Delta x^2)$, unconditionally stable, implicit

The advantage of Crank-Nicolson: we can take $\Delta t = \Delta x$ and still be as accurate as $\theta = 0$ or $\theta = 1$ method with $\Delta t = (\Delta x)^2$.

end of §2.11

Convergence of θ-method

To prove convergence, we use a maximum principle.

Maximum Principle: $\partial_t u = \partial_{xx} u$ satisfies a maximum principle: $u(x,t)$ is bounded by the extreme values attained by the initial condition and the values on the boundary.

The θ-method satisfies maximum principle but only if $\nu(1-\theta) \le \frac{1}{2}$.

Theorem 2.2: The θ-method with $0 \le \theta \le 1$ and $\nu(1-\theta) \le \frac{1}{2}$ yields $\{U_j^n\}$ satisfying $U_{min} \le U_j^n \le U_{max}$ where — (*)

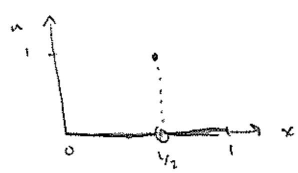
$$U_{min} = \min\{U_0^m,\ 0 \le m \le n;\ U_j^0,\ 0 < j \le J;\ U_J^m,\ 0 \le m \le n\}$$

$$U_{max} = \max\{U_0^m,\ 0 \le m \le n;\ U_j^0,\ 0 < j \le J;\ U_J^m,\ 0 \le m \le n\}.$$

If (*) is satisfied (stability) and the method is consistent then the method converges.

Example $\partial_t u = \partial_{xx} u \quad x \in [0,1]$

$$u(x,0) = \begin{cases} 1 & x = \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

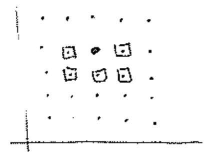


Proof: The θ-method is

$$(1+2\nu\theta)U_j^{n+1} = \theta\nu(U_{j-1}^{n+1}+U_{j+1}^{n+1}) + (1-\theta)\nu(U_{j-1}^{n}+U_{j+1}^{n}) + (1-2(1-\theta)\nu)U_j^n.$$

Using that $\nu(1-\theta)\le\frac{1}{2}$, all coefficients on the right side are positive and sum to $1+2\nu\theta$. Assume U has a maximum at an interval point j and that this maximum is U_j^{n+1}. Define

$$U^* = \max\{U_{j-1}^{n+1}, U_{j+1}^{n+1}, U_{j-1}^{n}, U_j^n, U_{j+1}^n\}$$



Now

$$(1+2\nu\theta)U_j^{n+1} \le 2\theta\nu U^* + 2(1-\theta)\nu U^* + (1-2(1-\theta)\nu)U^* = (1+2\nu\theta)U^*.$$

So $U_j^{n+1}\le U^*$. But we assumed $U_j^{n+1}\ge U^*$, so we must have $U_j^{n+1}=U^*$. Apply again to neighbouring point, etc., to show that max is on boundary or initial condition. ∎

2016 03 14

We've seen the θ-method is

- consistent ($T\to 0$, $\Delta x, \Delta t\to 0$)
- stable (depending on θ,ν)

What about convergence? We will use the maximum theorem.

Theorem: If $\nu(1-\theta)\le\frac{1}{2}$ then the θ-method is convergent.

Proof: Define the error to be $e_j^n = U_j^n - u_j^n$. The numerical solution satisfies

$$(1+2\theta\nu)U_j^{n+1} = \theta\nu(U_{j-1}^{n+1}+U_{j+1}^{n+1}) + (1-\theta)\nu(U_{j-1}^{n}+U_{j+1}^{n}) + (1-2(1-\theta)\nu)U_j^n.$$

The exact solution satisfies

$$(1+2\theta\nu)u_j^{n+1} = \theta\nu(u_{j-1}^{n+1}+u_{j+1}^{n+1}) + (1-\theta)\nu(u_{j-1}^{n}+u_{j+1}^{n}) + (1-2(1-\theta)\nu)u_j^n + \Delta t T_j^{n+1/2}.$$ $\leftarrow$?

Subtracting these equations from each other we get

$$(1+2\theta\nu)e_j^{n+1} = \theta\nu(e_{j-1}^{n+1}+e_{j+1}^{n+1}) + (1-\theta)\nu(e_{j-1}^{n}+e_{j+1}^{n}) + (1-2(1-\theta)\nu)e_j^n - \Delta t T_j^{n+1/2}.$$

This holds at $j=1,\dots,J-1$. For the initial and boundary conditions,

$$U_j^0 = u_j^0 \longrightarrow e_j^0 = 0$$
$$U_0^m = u_0^m,\ U_J^m = u_J^m \to e_0^m = 0,\ e_J^m = 0.$$

So no errors in initial and boundary conditions. Define

$$E^n = \max_{0\le j\le J}|e_j^n|,\quad T^{n+1/2} = \max_{1\le j\le J-1}|T_j^{n+1/2}|.$$

We will find a bound for the error:

$$|(1+2\theta\nu)e_j^{n+1}| = |\theta\nu(e_{j-1}^{n+1}+e_{j+1}^{n+1}) + (1-\theta)\nu(e_{j-1}^{n}+e_{j+1}^{n}) + (1-2(1-\theta)\nu)e_j^n - \Delta t\, T_j^{n+1/2}|$$

Since $\nu(1-\theta) \le \frac{1}{2}$, all coefficients are non-negative. So using the triangle inquality

$$(1+2\theta\nu)|e_j^{n+1}| \le \theta\nu(|e_{j-1}^{n+1}|+|e_{j+1}^{n+1}|) + (1-\theta)\nu(|e_{j-1}^{n}|+|e_{j+1}^{n}|) + (1-2(1-\theta)\nu)|e_j^n| + \Delta t\,|T_j^{n+1/2}|.$$

Hence

$$(1+2\theta\nu)E^{n+1} \le 2\theta\nu E^n + 2(1-\theta)\nu E^n + (1-2(1-\theta)\nu)E^n + \Delta t\, T^{n+1/2}$$

and so

$$E^{n+1} \le E^n + \Delta t\, T^{n+1/2}.$$

Therefore

$$E^n \le \Delta t \sum_{m=0}^{n-1} T^{m+1/2} \le n\,\Delta t \max_n T^{n+1/2}.$$

The θ-method is consistent, so $T \to 0$ as $\Delta x, \Delta t \to 0$. Therefore $E \to 0$ ∎

The maximum principle gives only a sufficient stability condition. For example, for Crank-Nicolson ($\theta = 1/2$), we showed that we have shown unconditional stability, but to satisfy the maximum principle, we need $\nu \le 1$.

Summary of θ-methods:

* stability: $0 \le \theta < \frac{1}{2} \longrightarrow \nu \le \frac{1}{2}(1-\theta)^{-1}$
  $\frac{1}{2} \le \theta \le 1 \longrightarrow \forall \nu$
* consistency: $\theta = \frac{1}{2} \longrightarrow T \sim \mathcal{O}(\Delta t^2), \mathcal{O}(\Delta x^2)$
  $\theta \ne \frac{1}{2} \longrightarrow T \sim \mathcal{O}(\Delta t), \mathcal{O}(\Delta x^2)$
* convergence: max principle only holds if $\nu(1-\theta) \le \frac{1}{2}$.

Special cases:

$\theta = 0 \longrightarrow$ explicit
$\theta = \frac{1}{2} \longrightarrow$ Crank-Nicolson
$\theta = 1 \longrightarrow$ fully implicit

Testing your implementation

Assume we want to solve the heat equation

$$\begin{cases} \partial_t u = \partial_{xx} u & 0 \le x \le 1,\ t > 0 \\ u(0,t) = u(1,t) = 0 & t > 0 \\ u(x,0) = u_0(x) & 0 \le x \le 1 \end{cases}$$

Assume we have implemented the θ-method. Our theory says the θ-method is stable, consistent, and so we expect convergence as $\Delta x, \Delta t \to 0$. In practice, you will implement the θ-method, but there will be bugs in your code. One way of testing your code is by the method of manufactored solutions.

A good way to test your code is by comparing your numerical solution to the exact solution. Many times you don't have the exact solution, so create one.

Assume you want $u(x,t) = \sin(2\pi x)\cos(2\pi t)$ to be the exact solution.

Then

$$\partial_t u = -2\pi \sin(2\pi x)\sin(2\pi t),$$
$$\partial_{xx} u = -4\pi^2 \sin(2\pi x)\cos(2\pi t).$$

Our modified problem becomes

$$\begin{cases} \partial_t u - \partial_{xx} u = 4\pi^2 \cos(2\pi t)\sin(2\pi x) - 2\pi \sin(2\pi t)\sin(2\pi x) \\ u(0,t) = u(1,t) = 0 \\ u(x,0) = \sin(2\pi x). \end{cases}$$

Since we know the exact solution, we can compute

$$E = \max\{|e_j^n|,\ (x_j, t_n) \in [0,1] \times [0, T_F]\}.$$

We will look at the error on succesively refined grids. The theory says $E \sim \mathcal{O}(\Delta x^{rate})$.

Define E_1 to be the error on grid 1 and let E_2 be the error on grid 2. 2016 03 15

$$E_1 = C(\Delta x_1)^{rate}$$
$$E_2 = C(\Delta x_2)^{rate}$$
$$\Rightarrow \frac{E_1}{E_2} = \left(\frac{\Delta x_1}{\Delta x_2}\right)^{rate} \Rightarrow rate = \frac{\ln(E_1/E_2)}{\ln(\Delta x_1/\Delta x_2)}$$

Explict ($\theta = 0$): theory says $E \sim \mathcal{O}(\Delta t) \sim \mathcal{O}(\Delta x^2)$ for r fixed

Implicit ($\theta = 1$): theory says $E \sim \mathcal{O}(\Delta t) \sim \mathcal{O}(\Delta x^2)$ for r fixed

$$E \sim \mathcal{O}(\Delta t) \sim \mathcal{O}(\Delta x) \quad \text{for } \Delta t = c\Delta x$$

Crank-Nicolson ($\theta = \frac{1}{2}$): theory says $E \sim O(\Delta t^2) \sim O(\Delta x^2)$

§2.12 A three-time level scheme

To solve the heat equation we've used

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{U_{j-1}^n - 2U_j^n + U_{j+1}^n}{(\Delta x)^2}.$$

Explicit, but $O(\Delta t)$. An improvement was Crank-Nicolson:

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

$O(\Delta t^2)$, but implicit.

Idea: use central differences in time and space to find $O(\Delta t^2)$ method that is explicit.

$$\frac{\partial u}{\partial t}(x_j, t_n) \approx \frac{U_j^{n+1} - U_j^{n-1}}{2\Delta t}.$$

We get the scheme

$$\frac{U_j^{n+1} - U_j^{n-1}}{2\Delta t} = \frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{(\Delta x)^2},$$

which is explicit. Let's calculate the truncation error.

$$T_j^n = \frac{u_j^{n+1} - u_j^{n-1}}{2\Delta t} + \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{(\Delta x)^2}.$$

Use Taylor series around (x_j, t_n):

$$u_j^{n+1} = u_j^n + \Delta t\, u_t + \tfrac{1}{2}(\Delta t)^2 u_{tt} + \tfrac{1}{6}(\Delta t)^3 u_{ttt} + \cdots$$

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

$$\Rightarrow u_j^{n+1} - u_j^{n-1} = 2\Delta t\, u_t + \tfrac{1}{3}(\Delta t)^3 u_{ttt} + \cdots$$

$$u_{j+1}^n - 2u_j^n + u_{j-1}^n = (\Delta x)^2 u_{xx} + \tfrac{1}{12}(\Delta x)^4 u_{xxxx} + \cdots \quad \text{(from before)}$$

Substituting into the truncation error:

$$T_j^n = \cancel{(u_t - u_{xx})}^{0} + \tfrac{1}{6}(\Delta t)^2 u_{ttt} - \tfrac{1}{12}(\Delta x)^2 u_{xxxx} + \cdots = O\left((\Delta t)^2 + (\Delta x)^2\right)$$

What about stability? Fourier symbol

$$U_j^n = \lambda^n \exp(ikj\Delta x).$$

Then

$$\frac{\lambda^{n+1}\exp(ikj\Delta x) - \lambda^{n-1}\exp(ikj\Delta x)}{2\Delta t} = \frac{\lambda^n}{(\Delta x)^2}\left(\exp(ik(j+1)\Delta x) - 2\exp(ikj\Delta x) + \exp(ik(j-1)\Delta x)\right)$$

$$\Rightarrow \quad \frac{\lambda - \lambda^{-1}}{2\Delta t} = \frac{1}{(\Delta x)^2}\left(\exp(ik\Delta x) - 2 + \exp(-ik\Delta x)\right)$$

$$\Rightarrow \quad \frac{\lambda - 1/\lambda}{2} = 2\nu\left(\cos(k\Delta x) - 1\right)$$

$$\Rightarrow \quad \lambda^2 + 8\sin^2\left(\tfrac{k\Delta x}{2}\right)\lambda - 1 = 0$$

Solving for λ yields

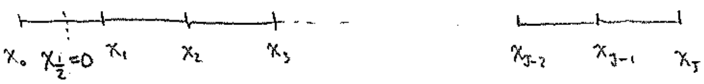
$$\lambda_{1,2} = -4\nu\sin^2\left(\frac{k\Delta x}{2}\right) \pm \sqrt{1 + 16\nu^2\sin^4\left(\frac{k\Delta x}{2}\right)}$$

For stability we need $|\lambda| < 1$, which we do not have

2016 03 17

Presentation about a failed discretization attempt.

$$\partial_x u = \alpha(t)u + g(t)$$



$$\partial_x u\Big|^n_{j=\frac{1}{2}} \approx \frac{U_1^n - U_0^n}{\Delta x} \qquad u^n_{1/2} \approx \frac{U_0^n + U_1^n}{2}$$

$$\frac{U_1^n - U_0^n}{\Delta x} = \alpha(t_n)\frac{U_0^n + U_1^n}{2} + g(t_n)$$

$$U_0^n = \frac{1 - \frac{1}{2}\alpha^n\Delta t}{1 + \frac{1}{2}\alpha^n\Delta x}U_1^n - \frac{\Delta x}{1 + \frac{1}{2}\alpha^n\Delta x}g^n$$

Truncation error: Taylor series expand around $x_{\frac{1}{2}} = \frac{1}{2}(x_0 + x_1)$

$$\frac{u_1^n - u_0^n}{\Delta x} - \frac{1}{2}\alpha^n(u_0^n + u_1^n) - g^n =$$

$$\frac{1}{\Delta x}\Bigg(\left(u + \tfrac{1}{2}\Delta x u_x + \tfrac{1}{8}\Delta x^2 u_{xx} + \tfrac{1}{48}\Delta x^3 u_{xxx} + \cdots\right)^n_{1/2}$$

$$- \left(u - \tfrac{1}{2}\Delta x u_x + \tfrac{1}{8}\Delta x^2 u_{xx} - \tfrac{1}{48}\Delta x^3 u_{xxx} + \cdots\right)^n_{1/2}\Bigg)$$

$$-\frac{1}{2}\alpha^n\Bigg(\left(u + \tfrac{1}{2}\Delta x u_x + \tfrac{1}{8}\Delta x^2 u_{xx} + \tfrac{1}{48}\Delta x^3 u_{xxx} + \cdots\right)^n_{1/2}$$

$$+ \left(u - \tfrac{1}{2}\Delta x u_x + \tfrac{1}{8}\Delta x^2 u_{xx} - \tfrac{1}{48}\Delta x^3 u_{xxx} + \cdots\right)^n_{1/2}\Bigg)$$

$$= \left(u_x + \tfrac{1}{24}\Delta x^2 u_{xxx} + \cdots - \alpha^n\left(u + \tfrac{1}{8}\Delta x^2 u_{xx} + \cdots\right)\right)^n_{1/2} - g^n$$

$$= \underbrace{(u_x - \alpha u - g)^n_{1/2}}_{=0} + \left(\frac{1}{24}\Delta x^2 u_{xxx} - \frac{1}{8}\alpha^n \Delta x^2 u_{xx} + \cdots\right)^n_{1/2}$$

So

$$T_1^{n+1/2} = \frac{U_1^{n+1} - U_1^n}{\Delta t} - \frac{U_2^n - 2U_1^n + U_0^n}{(\Delta x)^2} - \frac{1}{1+\frac{1}{2}\alpha^n \Delta x}\frac{1}{\Delta x}\left(\frac{U_1^n - U_0^n}{\Delta x} - \frac{1}{2}\alpha^n (U_0^n + U_1^n) - g^n\right)$$

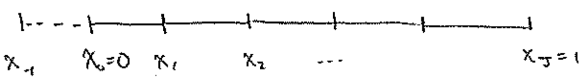
$$= \left(\frac{1}{2}\Delta t\, u_{tt} - \frac{1}{12}\Delta x^2 u_{xxxx} + \cdots\right)^n_1 - \frac{1}{1+\frac{1}{2}\alpha^n\Delta x}\left(\frac{1}{24}\Delta x\left(u_{xxx} - 3\alpha^n u_{xx}\right) + \cdots\right)^n_{1/2}$$

$$\sim O(\Delta x).$$

Thus we have consistency. One can check that for $\nu(1-\theta) \le \frac{1}{2}$ we satisfy the maximum principle.
Therefore we have convergence. It is also stable.

2016 03 21

Another option is to introduce a "ghost" node



So $\partial_x u(0,t) = \alpha(t) u(0,t) - g(t)$ is now approximated as

$$\frac{U_1^n - U_{-1}^n}{2\Delta x} = \alpha^n U_0^n + g^n$$

The θ-method applied to the heat equation at $j=0$ gives

$$0 = \frac{U_0^{n+1} - U_0^n}{\Delta t} - \frac{\theta}{(\Delta x)^2}\left(U_1^{n+1} - 2U_0^{n+1} + U_{-1}^{n+1}\right) - \frac{1-\theta}{(\Delta x)^2}\left(U_1^n - 2U_0^n + U_{-1}^n\right).$$

Now subtract

$$0 = \frac{2\theta}{\Delta x}\left(\frac{U_1^{n+1} - U_{-1}^{n+1}}{2\Delta x} - \alpha^{n+1}U^{n+1} - g^{n+1}\right) + \frac{2(1-\theta)}{\Delta x}\left(\frac{U_1^n - U_{-1}^n}{2\Delta x} - \alpha^n U_0^n - g^n\right)$$

to eliminate U_{-1}^n and U_{-1}^{n+1}. Now we can compute the truncation error:

$$\frac{2\theta}{\Delta x}\left(\frac{u_1^{n+1} - u_{-1}^{n+1}}{2\Delta x} - \alpha^{n+1}u_0^{n+1} - g^{n+1}\right) = \cdots = \theta\left(\frac{1}{3}\Delta x\, u_{xx} + \cdots\right)^{n+1}_0$$

$$\frac{2(1-\theta)}{\Delta x}\left(\frac{u_1^n - u_{-1}^n}{2\Delta x} - \alpha^n u_0^n - g^n\right) = \cdots = (1-\theta)\left(\frac{1}{3}\Delta x\, u_{xx} + \cdots\right)^n_0$$

Combining these with the truncation error of the θ-method, we find that $T \sim O(\Delta x)$. Hence we have consistent.
It turns out that stability follows from $\nu(1-\theta)(1+\alpha^n\Delta x) \le \frac{1}{2}$.

Apply the Maximum Principle

$$E^n \le n\Delta t \max_n T^{n+1/2}.$$

Thus we have convergence.

§2.14 Heat Conservation

$$\begin{cases} u_t = u_{xx}, & x \in [0,1],\ t>0 \\ u_x(0,t) = g_0(t) \\ u_x(1,t) = g_1(t) \end{cases}$$

The total heat in our system at time t is defined as

$$h(t) = \int_0^1 u(x,t)\,dx.$$

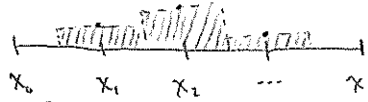
So by the heat equation,

$$\frac{dh}{dt} = \int_0^1 u_t\,dx = \int_0^1 u_{xx}\,dx = [u_x]_0^1 = g_1(t) - g_0(t).$$

We want our numerical method to imitate this property.

$$\frac{U_j^{n+1}-U_j^n}{\Delta t} = \frac{\theta}{(\Delta x)^2}\left(U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}\right) + \frac{1-\theta}{(\Delta x)^2}\underbrace{\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right)}_{=:M^n},$$

The total heat is $h(t) = \int_0^1 u(x,t)\,dt$.



$$H^n = \sum_{j=1}^{J-1} \Delta x\, U_j^n.$$

$$\frac{H^{n+1}-H^n}{\Delta t} = \int_0^1 \frac{U_j^{n+1}-U_j^n}{\Delta t}\,dx = \sum_{j=1}^{J-1}\int_{x_{j-1/2}}^{x_{j+1/2}} \frac{U_j^{n+1}-U_j^n}{\Delta t}\,dx = \sum_{j=1}^{J-1}\frac{\Delta x}{\Delta t}\left(U_j^{n+1}-U_j^n\right)$$

$$= \int_0^1 \left(\frac{\theta}{(\Delta x)^2}M^{n+1} + \frac{(1-\theta)}{(\Delta x)^2}M^{n+1}\right)dx = \frac{1}{\Delta x}\sum_{j=1}^{J-1}\left(\theta M^{n+1} + (1-\theta)M^n\right)$$

$$H^{n+1}-H^n = \frac{\Delta t}{\Delta x}\sum_{j=1}^{J-1}\left(\theta\left((U_{j+1}^{n+1}-U_j^{n+1}) - (U_j^{n+1}-U_{j-1}^{n+1})\right) + (1-\theta)\left((U_{j+1}^n - U_j^n) - (U_j^n - U_{j-1}^n)\right)\right)$$

$$= \frac{\Delta t}{\Delta x}\left(\theta(U_J^{n+1}-U_{J-1}^{n+1}) + \theta(U_J^n - U_{J-1}^n) - (1-\theta)(U_1^{n+1}-U_0^{n+1}) - (1-\theta)(U_1^n - U_0^n)\right)$$

since the sum is telescoping. The boundary conditions give

$$\frac{U_1^n - U_0^n}{\Delta x} = g_0^n \qquad \frac{U_J^n - U_{J-1}^n}{\Delta x} = g_1^n.$$

So

$$H^{n+1} - H^{n} = \Delta t\left(\Theta\left(g_1^{n+1} - g_0^{n+1}\right) - (1-\Theta)\left(g_1^{n} - g_0^{n}\right)\right).$$

This approximates

$$\frac{dh}{dt} = g_1 - g_0.$$

2016 03 22

Other PDEs

$$\partial_t u = b(x,t)\,\partial_{xx} u \qquad\qquad b(x,t) > 0$$

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = b_j^n \left(\frac{U_{j+1}^n - 2U_j^n + U_{j-1}^n}{(\Delta x)^2} \right)$$

Truncation error:

$$T(x,t) = \frac{1}{\Delta t}\left(u + \Delta t\, u_t + \tfrac{1}{2}(\Delta t)^2 u_{tt} + \cdots - u \right)$$

$$- \frac{b(x,t)}{(\Delta x)^2}\left(u + \Delta x\, u_x + \tfrac{1}{2}(\Delta x)^2 u_{xx} + \tfrac{1}{6}(\Delta x)^3 u_{xxx} + \cdots - 2u + u - \Delta x\, u_x + \tfrac{1}{2}(\Delta x)^2 u_{xx} - \tfrac{1}{6}(\Delta x)^3 u_{xxx} + \cdots \right)$$

$$= \underbrace{u_t - b(x,t) u_{xx}}_{=0} + \tfrac{1}{2}\Delta t\, u_{tt} - \tfrac{1}{12}(\Delta x)^2 b(x,t)\, u_{xxxx} + \cdots$$

So $T \to 0$ as $\Delta x, \Delta t \to 0$, so we have consistency.

Stability:

① $\partial_t u = \partial_{xx} u \longrightarrow U_j^{n+1} = U_j^n + \nu\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right)$

② $\partial_t u = b(x,t)\partial_{xx} u \rightarrow U_j^{n+1} = U_j^n + \nu b(x,t)\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right)$

① $\nu \le \frac{1}{2}$

② $\nu b_j^n \le \frac{1}{2} \quad \forall j, n$

For ①, the error bound was

$$E^n \le \tfrac{1}{2}\Delta t\left(M_{tt} + \tfrac{1}{6\nu} M_{xxxx}\right) T_F$$

For ②, we would find

$$E^n \le \tfrac{1}{2}\Delta t\left(M_{tt} + \tfrac{1}{6\nu} B M_{xxxx}\right) T_F$$

where $B = \sup(b)$.

θ-method:

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = \frac{1}{(\Delta x)^2} b(x_j, \underline{\underline{\ \ ?\ \ }})\left(\theta\left(U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}\right) + (1-\theta)\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right) \right)$$

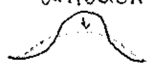
What should we take? t_n? t_{n+1}? $t_{n+1/2}$?

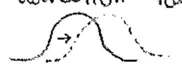
If we take $t_{n+1/2}$, then computing the truncation error and applying the maximum principle is almost the same as when $b = 1$.

If we take $\frac{1}{2}(b_j^{n+1} + b_j^n)$, then the analysis is much more complicated.

General linear parabolic PDE:

$$u_t = \underbrace{b(x,t)\,u_{xx}}_{\text{diffusion}} + \underbrace{a(x,t)\,u_x}_{\text{advection}} + \underbrace{c(x,t)\,u}_{\text{reaction term}} + \underbrace{d(x,t)}_{\text{source term}}, \qquad \begin{array}{l} b(x,t) > 0 \\ a(x,t) > 0 \end{array}$$





$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = b_j^n \frac{1}{(\Delta x)^2}\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right) + \frac{a_j^n}{2\Delta x}\left(U_{j+1}^n - U_{j-1}^n\right) + c_j^n U_j^n + d_j^n$$

The error is

$$e_j^{n+1} = (1 - 2\mu b_j^n + c_j^n \Delta t)\, e_j^n + (\mu b_j^n + \tfrac{1}{2}\nu a_j^n)\, e_{j+1}^n + (\mu b_j^n - \tfrac{1}{2}\nu a_j^n)\, e_{j-1}^n - \Delta t\, T_j^n$$

where

$$\mu = \frac{\Delta t}{(\Delta x)^2}, \qquad \nu = \frac{\Delta t}{\Delta x}.$$

For stability we need all coefficients to be nonnegative and sum no greater than unity:

i) $\quad 1 - 2\mu b_j^n + c_j^n \Delta t \geq 0 \longleftrightarrow 2\mu b_j^n - \Delta t\, c_j^n \leq 1$

ii) $\quad \mu b_j^n + \tfrac{1}{2}\nu a_j^n \geq 0 \longleftrightarrow -\tfrac{1}{2}\nu a_j^n \leq \mu b_j^n$

iii) $\quad \mu b_j^n - \tfrac{1}{2}\nu a_j^n \geq 0 \longleftrightarrow \tfrac{1}{2}\nu a_j^n \leq \mu b_j^n$

(ii) and (iii): $\quad \Delta x \leq \dfrac{2b_j^n}{|a_j^n|}$

iv) $\quad 1 + c_j^n \Delta t \leq 1 \longleftrightarrow c_j^n \leq 0$

In many applications, $|a_j^n| \gg b_j^n$, so this scheme is too restrictive.

Try

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} = b_j^n \frac{1}{(\Delta x)^2}\left(U_{j+1}^n - 2U_j^n + U_{j-1}^n\right) + \overbrace{\frac{a_j^n}{\Delta x}\left(U_{j+1}^n - U_j^n\right)}^{\text{changed}} + c_j^n U_j^n + d_j^n.$$

The error is

$$e_j^{n+1} = \underbrace{(1 - 2\mu b_j^n - a_j^n \nu)}_{\substack{\geq 0 \iff \\ \Delta t \leq \frac{(\Delta x)^2}{2b_j^n + \Delta x a_j^n}}} e_j^n + \underbrace{(\mu b_j^n + \nu a_j^n)}_{\substack{\geq 0 \\ \text{always}}} e_{j+1}^n + \underbrace{\mu b_j^n}_{\substack{\geq 0 \\ \text{always}}} e_{j-1}^n - \Delta t\, T_j^n$$

Next PDE:

$$u_t = \partial_x\left(p(x,t)\,\partial_x u\right)$$

Let $F = p(x,t)\,\partial_x u$. Then

$$\partial_x F = \frac{F_{j+1/2}^n - F_{j-1/2}^n}{\Delta x}$$

$$= \frac{P_{j+1/2}^n\, \partial_x u\big|_{j+1/2}^n - P_{j-1/2}^n\, \partial_x u\big|_{j-1/2}^n}{\Delta x}$$

Using

$$\partial_x U\Big|^n_{j+1/2} \approx \frac{U^n_{j+1} - U^n_j}{\Delta x}, \quad \partial_x U\Big|^n_{j-1/2} \approx \frac{U^n_j - U^n_{j-1}}{\Delta x},$$

we get

$$\partial_x F = \frac{1}{(\Delta x)^2}\left(P^n_{j+1/2}(U^n_{j+1} - U^n_j) - P^n_{j-1/2}(U^n_j - U^n_{j-1})\right).$$

2016 03 24

See presentation on learn.

2016 03 28
(written 03 29)

Consider the non-linear PDE

$$\begin{cases} \partial_t u = a(u)\,\partial_{xx} u, & a > 0 \\ u(0,t) = 0 \\ u(1,t) = 0 \\ u(x,0) = u_0(x) \end{cases}$$

Take the discretization

$$\frac{U^{n+1}_j - U^n_j}{\Delta t} = a(U^n_j)\,\frac{U^n_{j+1} - 2U^n_j + U^n_{j-1}}{(\Delta x)^2}.$$

To find the error, consider

$$U^{n+1}_j = U^n_j + \nu a(U^n_j)\,\frac{U^n_{j+1} - 2U^n_j + U^n_{j-1}}{\cancel{(\Delta x)^2}}, \qquad \text{— (1)}$$

$$u^{n+1}_j = u^n_j + \nu a(u^n_j)\,\frac{u^n_{j+1} - 2u^n_j + u^n_{j-1}}{\cancel{(\Delta x)^2}} + \Delta t\, T^n_j. \qquad \text{— (2)}$$

Note $a(U^n_j) \neq a(u^n_j)$ so we cannot combine terms. We Taylor expand:

$$\begin{aligned} a(u^n_j) &= a(U^n_j) + (u^n_j - U^n_j)\,\partial_u a(\eta) \\ &= a(U^n_j) - e^n_j q^n_j \end{aligned}$$

where $q^n_j = \partial_u a(\eta)$, for some η between U^n_j and u^n_j. Therefore

$$u^{n+1}_j = u^n_j + \nu a(U^n_j)(u^n_{j+1} - 2u^n_j + u^n_{j-1}) - \nu e^n_j q^n_j (u^n_{j+1} - 2u^n_j + u^n_{j-1}) - \Delta t\, T^n_j \qquad \text{— (3)}$$

Now subtract (3) from (1):

$$e^{n+1}_j = e^n_j + \nu a(U^n_j)(e^n_{j+1} - 2e^n_j + e^n_{j-1}) + \nu e^n_j q^n_j \underbrace{(u^n_{j+1} - 2u^n_j + u^n_{j-1})}_{(\Delta x)^2 \left(u_{xx} + \frac{1}{12}(\Delta x)^2 u_{xxxx} + \cdots\right)\Big|^n_j} - \Delta t\, T^n_j$$

We need $1 - 2\nu a(U^n_j) \geq 0$ $\forall j$, and so

$$\Delta t \leq \frac{\frac{1}{2}(\Delta x)^2}{\max_j a(U^n_j)}.$$

Note that this means that in your code, you need to recompute timestep each time.
Assume the solution is smooth enough such that

$$|u_{j+1}^n - 2u_j^n + u_{j-1}^n| \le (\Delta x)^2 M_{xx}$$

(i.e. Taylor series expansion bounded). Also assume that q_j^n is well-behaved:

$$|q_j^n| \le K.$$

Given these assumptions, we bound the error:

$$|e_j^{n+1}| = |(1-2\nu a(U_j^n))e_j^n + \nu a(U_j^n)e_{j+1}^n + \nu a(U_j^n)e_{j-1}^n + \nu e_j^n q_j^n (u_{j+1}^n - 2u_j^n + u_{j-1}^n) - \Delta t T_j^n|$$
$$\le (1-2\nu a(U_j^n))|e_j^n| + \nu a(U_j^n)|e_{j+1}^n| + \nu a(U_j^n)|e_{j-1}^n| + \nu K(\Delta x)^2 M_{xx}|e_j^n| + \Delta t \overline{T}$$

where $\overline{T} = \max|T_j^n|$. Introducing $E^n = \max_j |e_j^n|$, we have

$$|E^{n+1}| \le (1-2\nu a(U_j^n) + \nu a(U_j^n) + \nu a(U_j^n))E^n + \nu K(\Delta x)^2 M_{xx} E^n + \Delta t \overline{T}$$
$$= (1 + K M_{xx}\Delta t)E^n + \Delta t + \overline{T}.$$

By induction, like in the linear case, we can show

$$E^{n+1} \le (1+KM_{xx}\Delta t)^n \Delta t \overline{T} + n\Delta t \overline{T}.$$

Note

$$(1+KM_{xx}\Delta t)^n \le \exp(KM_{xx} n \Delta t) \le \exp(KM_{xx}T_F),$$

and so

$$E^{n+1} \le \Delta t \exp(KM_{xx}T_F)\overline{T} + T_F\overline{T}.$$

However this bound is large and not useful. It only shows stability.
So for explicit method, it's easy to implement the non-linear if you've already implemented the linear.
What would we expect and implicit method to look like?

$$U_j^{n+1} = U_j^n + \nu a(U_j^{n+1})(U_{j+1}^{n+1} - 2U_j^{n+1} + U_{j-1}^{n+1}).$$

We can do as before, and set up a matrix system. This will be of the form $A(U^{n+1})U^{n+1} = F$. How do we solve this?

Fixed Point Iteration: Assume we want to solve $G(x) = x$. Algorithm: Starting with x^0, set $x^{k+1} = G(x^k)$. Then $x^0, x^1, x^2, \ldots$ converges to x such that $G(x) = x$.

So for solving $A(U^{n+1})U^{n+1} = F$, we have $U^{n+1} = G(U^{n+1}) := A^{-1}(U^{n+1})F$.
We start with $U^{n+1,0} := U^n$, and take $U^{n+1,k+1} = A^{-1}(U^{n+1,k})F$. Note that even though we don't know $A^{-1}(U^{n+1})F$, we do know $A^{-1}(U^{n+1,k})$.
So once you go into non-linearity, things get harder. You need to keep track of many non-linear terms, convergence of systems.