

# Math 228B Lec 2

last time: overview (PDE 200), heat equation on  $[0, \pi]$   
 heat equation on  $\mathbb{R}$

today: finite difference notation, truncation error

Question: why use finite differences (or any numerical method) if we have an exact formula for the solution?

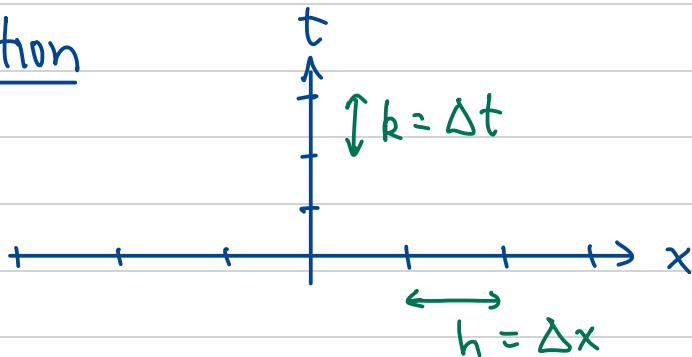
$$u(x, t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-\frac{(x-\xi)^2}{4t}} g(\xi) d\xi \quad (\text{real line version})$$

1. We would have to compute this integral somehow (numerics are still involved in evaluating the exact solution)
2. These exact solutions do not generalize to more complicated geometries
3. It's useful to develop intuition about the numerical schemes on problems that you completely understand

Finite difference notation

discretization:

(uniform grid)



numerical solution:  $u_j^n \approx u(x_j, t_n)$  exact solution  
 space:  $x_j = jh$  time:  $t_n = nk$

for the initial condition, set  $u_j^0 = g(jh)$  sample the exact initial condition

Simplest scheme for  $u_t = u_{xx}$  : forward time, centered space

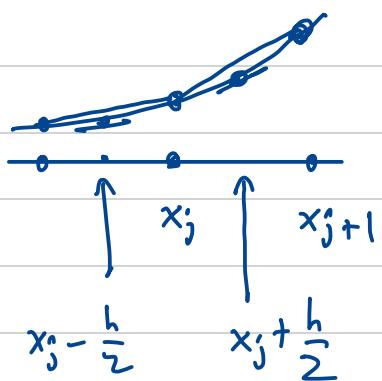
$$u_t \approx \frac{u(x_i, t_n + k) - u(x_i, t_n)}{k} \quad k = \Delta t$$

$$\approx \frac{1}{k} (u_j^{n+1} - u_j^n) = D_t^+ u_j^n$$

$$u_{xx} \approx \frac{u_x(x_i + \frac{h}{2}, t_n) - u_x(x_i - \frac{h}{2}, t_n)}{h} \quad h = \Delta x$$

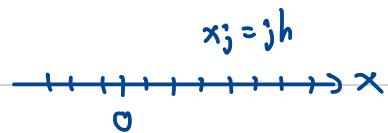
$$\approx \frac{1}{h} \left[ \frac{u_{j+1}^n - u_j^n}{h} - \frac{u_j^n - u_{j-1}^n}{h} \right] = \frac{1}{h} (D_x^+ u_j^n - D_x^- u_j^n)$$

$$= \frac{1}{h^2} (u_{j+1}^n - 2u_j^n + u_{j-1}^n) = D_x^+ D_x^- u_j^n$$



secant line  
 approximations of  
 the slopes at the  
 midpoints

setup:  $f(x)$  defined for  $x \in \mathbb{R}$



Convention 1:  $f_j = f(x_j)$  are sampled values,  $x_j = jh$  (uniform grid)

Convention 2:  $f_j$  is a discrete sequence (e.g. the numerical solution)

and we are hoping to achieve  $f_j \approx f(x_j)$  e.g. the exact solution

define  $D^+ f_j = \frac{f_{j+1} - f_j}{h}$ ,  $D^- f_j = \frac{f_j - f_{j-1}}{h}$

$$D^0 f_j = \frac{f_{j+1} - f_{j-1}}{2h}, \quad D^+ D^- f_j = \frac{f_{j+1} - 2f_j + f_{j-1}}{h^2}$$

$f$  can mean  $f(x)$  or  $\{f_j\}_{j \in \mathbb{Z}}$ ,  $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$   
integers

$D^+ f$  can mean  $\underbrace{D^+ f}_\text{a new function}(x) = \frac{f(x+h) - f(x)}{h}$  continuous version

or  $D^+ f_j = \underbrace{(D^+ f)}_\text{a new sequence}_j = \frac{f_{j+1} - f_j}{h}$  sequence version  
(more common)

$$D^+ D^- f_j = (D^+ (D^- f))_j = \frac{1}{h} \left[ (D^- f)_{j+1} - (D^- f)_j \right]$$

$$= \frac{1}{h} \left[ \frac{f_{j+1} - f_j}{h} - \frac{f_j - f_{j-1}}{h} \right]$$

↑  
j+1 term of the sequence  $D^- f$

FTCS scheme for  $u_t = u_{xx}$ :  $D_t^+ u = D_x^+ D_x^- u$

$$\frac{1}{k} (u_j^{n+1} - u_j^n) = \frac{1}{h^2} (u_{j+1}^n - 2u_j^n + u_{j-1}^n)$$

solve for  $u_j^{n+1}$ , set  $\nu = \frac{k}{h^2}$

$$u_j^{n+1} = \nu u_{j+1}^n + (1-2\nu) u_j^n + \nu u_{j-1}^n$$

Scheme: method of advancing the numerical solution  
from  $t_n$  to  $t_{n+1}$

truncation error: what's left over when you plug  
the exact solution  $u(x, t)$  into the scheme

$$(\star) \quad D_t^+ u(x_j, t_n) = D_x^+ D_x^- u(x_j, t_n) + \tau_j^n$$

note: in 228A, we include a factor of  $k$  in  $\tau_n$

$y' = f(y)$ , Euler's method:

$$y(t_n + k) = y(t_n) + k f(y(t_n)) + \tau_n^{228A}$$

$$\frac{y(t_n + k) - y(t_n)}{k} = f(y(t_n)) + \tau_n^{228B}$$

$$\tau_n^{228A} = k \tau_n^{228B}$$

In  $(\star)$ ,  $u(x_j, t_n)$  denotes the exact solution while  $u_j^n$  denotes the numerical solution, so  $u_j^n \approx u(x_j, t_n)$

↑  
not exactly equal

convention 2      convention 1

In other contexts, (approximation theory, Fourier analysis)

$u_j^n$  is defined by sampling, so  $u_j^n = u(x_j, t_n)$

and we would switch to  $U_j^n$  (capital letter)

or  $w_j^n$  for the numerical solution.

### Taylor expansions

$$D_t^+ u(x_j, t_n) = \underbrace{D_x^+ D_x^- u(x_j, t_n)}_A + \underbrace{\tau_j^n}_B$$

$$A = \frac{u(x_j, t_n+k) - u(x_j, t_n)}{k}$$

$$= \frac{u + k u_t + \frac{k^2}{2} u_{tt} + \frac{k^3}{6} u_{ttt}(x_j, t_n + \theta_1 k) - u}{k}$$

$\theta_1 \in (0, 1)$  from Taylor's theorem with remainder.

Without arguments,  $u, u_t, u_{tt}$  are evaluated at  $(x_j, t_n)$

$$A = u_t + \frac{k}{2} u_{tt} + \frac{k^2}{6} u_{ttt}(x_j, t_n + \theta_1 k) \quad \text{exact formula}$$

$$B = \frac{1}{h^2} \left[ u(x_j + h, t_n) - 2u(x_j, t_n) + u(x_j - h, t_n) \right]$$

$$= \frac{1}{h^2} \begin{bmatrix} u + hu_x + \frac{h^2}{2} u_{xx} + \frac{h^3}{6} u_{xxx} + \dots \\ -2u \\ + u - hu_x + \frac{h^2}{2} u_{xx} - \frac{h^3}{6} u_{xxx} + \dots \end{bmatrix}$$

$$= u_{xx} + \frac{h^2}{12} u_{xxxx} + \frac{h^4}{720} \left( \begin{array}{l} \partial_x^6 u(x_j + \theta_2 h, t_n) \\ + \partial_x^6 u(x_j - \theta_3 h, t_n) \end{array} \right)$$

$\uparrow$   
exact formula

$$\partial_x^6 u = u_{xxxxxx}$$

$$\tau_j^n = A - B = (u_t - u_{xx}) + \frac{1}{2} \left( k u_{tt} - \frac{h^2}{6} u_{xxxx} \right) + O(k^2 + h^4)$$

remainder terms  
in Taylor's theorem

Since  $u(x, t)$  is the exact solution,  $u_t(x_j, t_n) = u_{xx}(x_j, t_n)$

$$\text{and } u_{tt} = (u_t)_t = (u_{xx})_t = (u_t)_{xx} = u_{xxxx}$$

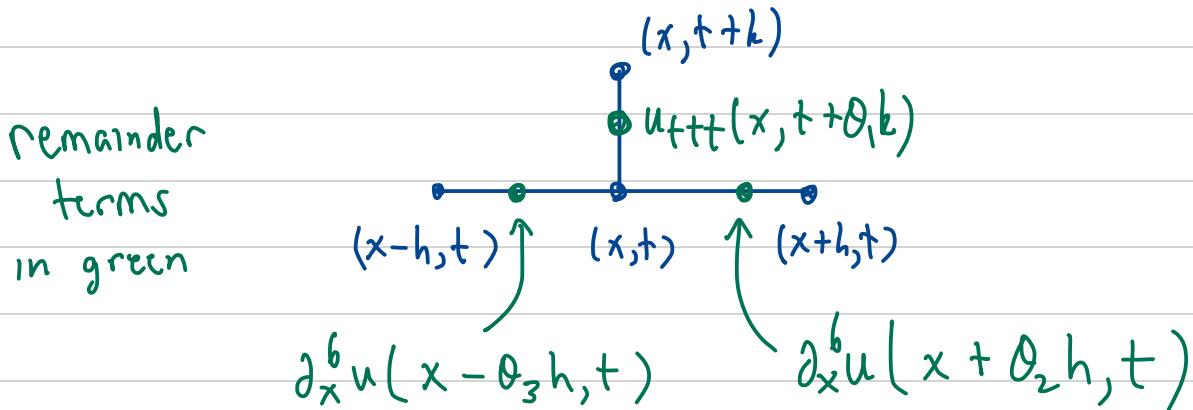
$$\text{so } \tau_j^n = \frac{h^2}{2} \left( v - \frac{1}{6} \right) u_{xxxx} + O(k^2 + h^4), \quad v = \frac{k}{h^2}$$

$$\text{If } v = \frac{1}{6}, \quad |\tau_j^n| \leq M \left( \frac{k^2}{6} + \frac{h^4}{360} \right) = \frac{M h^4}{135}$$

$\uparrow$   
 $k = h^2/6$

$$M = \max_{x,t} |u_{ttt}(x,t)| = \max_{x,t} |\partial_x^6 u(x,t)|$$

where the max is over the stencil



Since we want a bound on all the  $\tau_j^n$ 's, we take the max over  $x \in \mathbb{R}$  and  $0 \leq t \leq T$  = final time

One can show from the exact solution formula that

$$M \leq \max_{x \in \mathbb{R}} |\partial_x^6 g(x)|, \quad g = \text{initial condition}$$

If  $\nu \neq \frac{1}{6}$ , we could have stopped sooner in the Taylor's theorem with remainder formulas. Result:

$$\tau_j^n = A - B = u_t - u_{xx} + O(h^2)$$

$$|\tau_j^n| \leq M \left( \frac{k}{2} + \frac{h^2}{12} \right) = \frac{M}{2} \left( \nu + \frac{1}{6} \right) h^2$$

different M

$$M = \max_{(x,t) \in \text{stencil}} |u_{ttt}| = \max_{(x,t) \in \text{stencil}} |u_{xxxx}| \leq \max_{x \in \mathbb{R}} |\partial_x^4 g(x)|$$