COMPUTATIONAL PDE LECTURE 2

LUCAS BOUCK

1. Outline of today

- Derive Poisson's equation in 3D using divergence theorem.
- Begin discussion of second order elliptic boundary value problems (BVP) in 1D
  - Green's function

2. Derivation of Poisson's Equation from Electrostatics

Here, we derive Poisson's equation in 3D using divergence theorem. The key takeaway from this section is the use of divergence theorem for PDEs. Divergence theorem is a critical tool that we will use throughout the course.

We first begin with two facts from electrostatics:

- Let $\rho : \mathbb{R}^3 \to \mathbb{R}$ be a charge density. We also consider $\Omega \subset \mathbb{R}^3$ to be an open, bounded, and simply connected domain. *Gauss's Law* is an experimental law that states that the electric field $\mathbf{e} : \mathbb{R}^3 \to \mathbb{R}^3$ that is produced by ρ must satisfy

$$\int_{\partial\Omega} \mathbf{e}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) dS = \int_{\Omega} \underbrace{\frac{\rho(\mathbf{x})}{\varepsilon_0}}_{=:f(\mathbf{x})} d\mathbf{x} \tag{1}$$

E field dotted w/ unknown

  where
  - $\partial\Omega$ denotes the boundary of Ω, which we will assume is smooth.
  - $\mathbf{n} : \partial\Omega \to \mathbb{R}^3$ is the outward unit normal vector of $\partial\Omega$
  - ε_0 is the electric permitivity of free space
- An electrostatic field $\mathbf{e} : \mathbb{R}^3 \to \mathbb{R}^3$ is conservative. A fact that follows from $\mathbf{e}$ being conservative is there exists a potential $u : \mathbb{R}^3 \to \mathbb{R}$ such that $\nabla u = \mathbf{e}$.

Remark 2.1 (vector notation)**.** Bold face $\mathbf{x}$ denotes a vector $\mathbf{x} = (x, y, z)$ and boldface $\mathbf{v} = (v_1, v_2, v_3)$.

Remark 2.2 (volume and surface integrals)**.** The integral $\int_\Omega f(\mathbf{x}) d\mathbf{x}$ can be written in more familiar multivariable calculus language as

$$\int_\Omega f(\mathbf{x}) d\mathbf{x} = \iiint_\Omega f(x, y, z) dx dy dz,$$

surface integrals :

$\int_\Omega f(\vec{x})\, d\vec{x} = \iiint_\Omega f(x,y,z)\, dx dy dz$

Date: August 30, 2023.

and a surface integral $\int_{\partial\Omega} f(\mathbf{x})dS$ can be written in more familiar multivariable calculus language as

$$\int_{\partial\Omega} f(\mathbf{x})dS = \iint_{\partial\Omega} f(x,y,z)dS,$$

To derive Poisson's equation, we need another tool, which is the divergence theorem. We first define divergence below.

Definition 2.1 (divergence)**.** The divergence of $\mathbf{v} = (v_1, v_2, v_3)$ is

$$\operatorname{div} \mathbf{v}(\mathbf{x}) = \nabla \cdot \mathbf{v}(\mathbf{x}) = \frac{\partial}{\partial x} v_1(\mathbf{x}) + \frac{\partial}{\partial y} v_2(\mathbf{x}) + \frac{\partial}{\partial z} v_3(\mathbf{x}) \tag{2}$$

If $\mathbf{v} : \mathbb{R}^2 \to \mathbb{R}^2$ is a 2D vector field, then the divergence is

$$\operatorname{div} \mathbf{v}(\mathbf{x}) = \nabla \cdot \mathbf{v}(\mathbf{x}) = \frac{\partial}{\partial x} v_1(\mathbf{x}) + \frac{\partial}{\partial y} v_2(\mathbf{x}).$$

Theorem 2.1 (divergence theorem)**.** Let $\Omega \subset \mathbb{R}^3$ be an open, bounded domain with smooth boundary, and let $\mathbf{n} : \partial\Omega \to \mathbb{R}^3$ be the outward unit normal of Ω. Then for any continuously differentiable vector field $\mathbf{v} : \overline{\Omega} \to \mathbb{R}^3$, we have

$$\int_\Omega \operatorname{div} \mathbf{v}(\mathbf{x})d\mathbf{x} = \int_{\partial\Omega} \mathbf{v}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})dS \tag{3}$$

Remark 2.3 (divergence theorem with dimension 1 and 2)**.** The divergence theorem in 2D is

$$\int_\Omega \operatorname{div} \mathbf{v}(\mathbf{x})d\mathbf{x} = \int_{\partial\Omega} \mathbf{v}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})ds$$

where $\int_{\partial\Omega} \mathbf{v}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})ds$ is the path integral over the curve $\partial\Omega$. The divergence theorem in 1D is

$$\int_a^b \frac{dv}{dx}(x)dx = v(b) - v(a), \tag{4}$$

which is the Fundamental Theorem of Calculus.

We return to the derivation and now combine the divergence theorem (3) with Gauss's law (1) to obtain

$$\int_\Omega \operatorname{div} \mathbf{e}(\mathbf{x})d\mathbf{x} = \int_\Omega f(\mathbf{x})d\mathbf{x}.$$

We substitute $\mathbf{e} = \nabla u$ (since $\mathbf{e}$ is conservative) in the above equation to obtain

$$\int_\Omega \operatorname{div} \nabla u(\mathbf{x})d\mathbf{x} = \int_\Omega f(\mathbf{x})d\mathbf{x}.$$

Using the definition of divergence (2), we have

$$\operatorname{div} \nabla u(\mathbf{x}) = \frac{\partial}{\partial x}\frac{\partial}{\partial x}u(\mathbf{x}) + \frac{\partial}{\partial y}\frac{\partial}{\partial y}u(\mathbf{x}) + \frac{\partial}{\partial z}\frac{\partial}{\partial z}u(\mathbf{x}) = u_{xx}(\mathbf{x}) + u_{yy}(\mathbf{x}) + u_{zz}(\mathbf{x})$$

Definition 2.2 (Laplacian of u)**.** We often denote

$$\Delta u(\mathbf{x}) = u_{xx}(\mathbf{x}) + u_{yy}(\mathbf{x}) + u_{zz}(\mathbf{x})$$

and call Δu the Laplacian of u.

We finally insert $\Delta u(\mathbf{x}) = \operatorname{div} \nabla u(\mathbf{x})$ into the above equation to obtain

$$\int_\Omega \Delta u(\mathbf{x}) d\mathbf{x} = \int_\Omega f(\mathbf{x}) d\mathbf{x}.$$

Note that Ω has been an arbitrary domain with smooth boundary. In particular, we can take $\Omega = B_r(\mathbf{x}_0)$ to be a ball of radius r centered at x_0. Dividing both sides of the above equation by the volume of $B_r(\mathbf{x}_0)$, which is $\frac{4}{3}\pi r^3$, we have

$$\frac{1}{\frac{4}{3}\pi r^3}\int_{B_r(\mathbf{x}_0)} \Delta u(\mathbf{x}) d\mathbf{x} = \frac{1}{\frac{4}{3}\pi r^3}\int_{B_r(\mathbf{x}_0)} f(\mathbf{x}) d\mathbf{x}.$$

Taking the limit as $r \to 0$ of both sides yields

$$\Delta u(\mathbf{x}_0) = f(\mathbf{x}_0),$$

which is Poisson's equation.

Exercise 2.1 (optional)**.** Prove the following for a continuous function $f : \mathbb{R}^3 \to \mathbb{R}$:

$$\lim_{r\to 0} \frac{1}{\frac{4}{3}\pi r^3}\int_{B_r(\mathbf{x}_0)} f(\mathbf{x}) d\mathbf{x} = f(\mathbf{x}_0).$$

3. Elliptic Boundary Value Problems in 1D

We'll focus on properties of Poisson's equation and similar problems in 1 dimension and use

$$\begin{cases} -u''(x) = f(x) & \text{for } x \in (0,1), \\ u(0) = u(1) = 0 \end{cases} \tag{5}$$

as the example. Note that we are enforcing

$$u(0) = u(1) = 0$$

rather than enforcing the initial conditions

$$u(0) = u_0 \text{ and } u'(0) = v_0.$$

The condition $u(0) = u(1) = 0$ is known as a **boundary condition**. The combination of a PDE and boundary conditions is called a **boundary value problem (BVP)** . There are different types of boundary conditions.

- Dirichlet boundary conditions: We set $u(0) = u_l$ and $u(1) = u_r$
- Neumann boundary conditions: We set $u'(0) = g_l$ and $u'(1) = g_r$
- Robin boundary conditions: We set $\alpha u(0) + \beta u'(0) = b_l$ and $\gamma u(1) + \delta u'(1) = b_r$

We may also mix these boundary conditions. For example, we may have Dirichlet on the left and Neumann on the right. The particular boundary conditions we are studying are also called **homogenous** boundary conditions, which means $u \equiv 0$ satisfies the boundary condition. A PDE is called homogenous if $u \equiv 0$ solves the PDE. The differential equation $-u'' = f$ is homogenous if and only if $f(x) = 0$ for all $x \in (0, 1)$.

A boundary value problem will translate better to PDEs in higher dimensions.

Remark 3.1. The minus sign in front of u'' is a common convention.

3.1. **Constructing Solutions: Green's Functions.** This discussion follows Chapter 2.1 of the textbook.

We will now actually construct an exact solution to (5) of the form

$$u(x) = \int_0^1 G(x, y) f(y) dy,$$

where G is known as a **Green's function**. If we can find such a function G, then this gives a very general way of constructing solutions to (5) for any f.

We'll construct G by reverse engineering. Suppose u is a twice continuously differentiable solution of (5). First we apply Fundamental Theorem of Calculus (4) to u to get

$$u(x) = u(0) + \int_0^x u'(y) dy = \int_0^x u'(y) dy.$$

Note that we used $u(0) = 0$ from (5). Applying Fundamental Theorem of Calculus (4) again to $u'(y) = u'(0) + \int_0^y u''(s) ds$, we have

$$u(x) = \int_0^x \left[u'(0) + \int_0^y u''(s) ds \right] dy = xu'(0) + \int_0^x \int_0^y u''(s) ds dy$$

Note that we now have u'' in the above expression for u. Using the differential equation $-u''(x) = f(x)$ in (5), we write

$$u(x) = xu'(0) - \int_0^x \int_0^y f(s) ds dy.$$

To simplify the next step, we write define

$$F(y) := \int_0^y f(s)ds,$$

and simplify the expression for u:

$$u(x) = xu'(0) - \int_0^x F(y)dy. \tag{6}$$

Note that $u(0) = 0$. In order to proceed to simplify the farthest right term of (6), we'll need integration by parts.

Lemma 3.1 (integration by parts)**.** Let $u, v : [a, b] \to \mathbb{R}$ be continuously differentiable on (a, b), then

$$\int_a^b w'(x)v(x)dx = w(b)v(b) - w(a)v(a) - \int_a^b w(x)v'(x)dx \tag{7}$$

Exercise 3.1. Prove integration by parts using Fundamental Theorem of Calculus (4) and the product rule: $(wv)'(x) = w'(x)v(x) + w(x)v'(x)$.

We now apply (7) to the farthest right term of (6) with $v(y) = F(y)$ and $w(y) = y$, $a = 0$ and $b = x$ to get

$$\int_0^x F(y)dy = \int_0^x F(y) \cdot 1dy = xF(x) - \cancel{0 \cdot F(0)}^{0} - \int_0^x yF'(y)dy$$

We then use the definition of $F(y) = \int_0^y f(s)ds$ to get $F'(y) = f(y)$ and

$$\int_0^x F(y)dy = x\int_0^x f(s)ds - \int_0^x yf(y)dy = \int_0^x (x - y)f(y)dy$$

Thus (6) becomes

$$u(x) = xu'(0) - \int_0^x (x - y)f(y)dy. \tag{8}$$

Substituting $x = 1$ into (8) and using the boundary condition $u(1) = 0$, we have

$$0 = u'(0) - \int_0^1 (1 - y)f(y)dy.$$

Hence $u'(0) = \int_0^1 (1 - y)f(y)dy$ and (8) simplifies to

$$u(x) = x\int_0^1 (1 - y)f(y)dy - \int_0^x (x - y)f(y)dy$$

Note that

$$\int_0^x (x - y)f(y)dy = \int_0^x (x - y)f(y)dy + \int_x^1 0 \cdot f(y)dy = \int_0^1 \max\{(x - y), 0\}f(y)dy,$$

so the u may be expressed as

$$u(x) = \int_0^1 \left[x(1-y) - \max\{(x-y), 0\}\right] f(y)dy.$$

Hence, we may write the solution u as:

$$u(x) = \int_0^1 G(x,y)f(y)dy, \tag{9}$$

where

$$G(x,y) = \left[x(1-y) - \max\{(x-y), 0\}\right] = \begin{cases} x(1-y), & y > x \\ y(1-x) & y \leq x \end{cases}.$$

This derivation is reverse engineering. Next class, we will show that the formula (9) is a an actual solution to (5).

Green's Function Process :

$u(x) = \int_0^1 G(x,y) f(y)dy$

Construct solution in form $u(x) = \int_0^1 G(x,y) f(y)dy$

1. apply FTOC to $u(x)$
2. apply FTOC to $u'(x)$

} can apply BC's in this process

apply FTOC twice, applying any relevant BC's

3. use above + $-u'' = f$ to write expression for $u(x)$ — IBP to simplify
4. define $F(y) = \int_0^y f(s)ds$ and use IBP to simplify $\int_0^x F(y)dy$ — $\int_0^x F(y)dy$
5. apply remaining BC's — remaining BC's
6. simplify the found form of $u(x)$ into compact $u(x) = \int_0^1 G(x,y) f(y)dy$

simplify into compact $u(x)$

COMPUTATIONAL PDE LECTURE 3

LUCAS BOUCK

1. Outline of today

- Finish discussion of Green's functions
- Maximum principle

2. Green's Functions

Recall we have been studying Poisson's equation 1 dimension:

$$\begin{cases} -u''(x) = f(x) & \text{for } x \in (0,1), \\ u(0) = u(1) = 0 \end{cases}. \tag{1}$$

We also showed that if u solves (1), then u takes the form

$$u(x) = \int_0^1 G(x,y) f(y) dy \tag{2}$$

where

$$G(x,y) = \big[x(1-y) - \max\{(x-y), 0\}\big] = \begin{cases} x(1-y), & x < y \\ y(1-x) & y \le x \end{cases} \tag{3}$$

is known as a Green's function.

Note that our derivation assumed we had a solution of (1) to start. We now will double check that (2) is indeed a solution of (1).

A useful tool for showing this is **Leibniz integral rule**.

Lemma 2.1 (Leibniz integral rule)**.** Let $a, b : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuously differentiable. Then

$$\frac{d}{dx} \int_{a(x)}^{b(x)} g(x,y) dy = g(x, b(x)) b'(x) - g(x, a(x)) a'(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} g(x,y) dy. \tag{4}$$

Date: September 1, 2023.

Remark 2.1 (special cases of Leibniz integral rule)**.** One special case of this rule is a version of fundamental theorem of calculus. If a is constant $b(x) = x$ and g does not depend on x, then

$$\frac{d}{dx}\int_a^x g(y)dy = g(x),$$

Another special case of Leibniz integral rule is if a, b do not depend on x, then,

$$\frac{d}{dx}\int_a^b g(x,y)dy = \int_a^b \frac{\partial}{\partial x}g(x,y)dy, \tag{5}$$

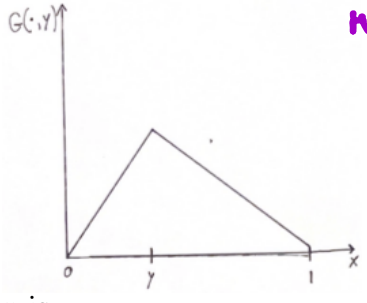
which we'll use when we study the heat equation.

We'll be unrigorous and apply (5) twice to u in (2) and compute:

$$u'(x) = \int_0^1 \frac{\partial}{\partial x}G(x,y)f(y)dy \quad \text{and} \quad u''(x) = \int_0^1 \frac{\partial^2}{\partial x^2}G(x,y)f(y)dy$$

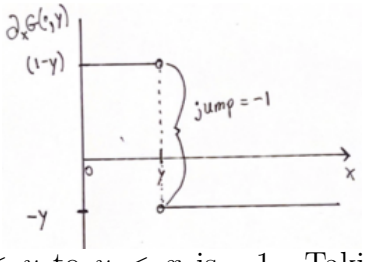
Recall that G is

$$G(x,y) = \begin{cases} x(1-y), & x < y \\ y(1-x) & y \le x \end{cases}.$$



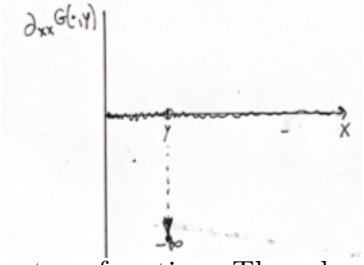
The partial derivative of G with respect to x is

$$\frac{\partial}{\partial x}G(x,y) = \begin{cases} (1-y) & x < y \\ \text{undefined} & y = x \\ y & y < x \end{cases}.$$



Notice that the jump from in $\partial_x G$ from $x < y$ to $y < x$ is -1. Taking another derivative G with respect to x is

$$\frac{\partial^2}{\partial x^2}G(x,y) = -\delta(x-y) = \begin{cases} 0 & x < y \\ -\infty & y = x \\ 0 & y < x \end{cases}.$$



Here, δ is a **Dirac delta function** and is not a true function. Though we define its

note the second derivative is dirac delta, from which we know the sifting property

integral with a continuous function f as

sifting: $\int_{-\infty}^{\infty} \delta(x-y)f(y)dy = f(x)$

$$\int_0^1 \delta(x-y)f(y)dy = f(x).$$

We then have

$$-u''(x) = -\int_0^1 \underbrace{\frac{\partial^2}{\partial x^2}G(x,y)}_{=-\delta(x-y)} f(y)dy = f(x).$$

This can all be made more rigorous by splitting the integral of

$$u(x) = \int_0^1 G(x,y)f(y)dy$$

can be made more rigorous by splitting G(x,y) into two integrals and apply leibniz to each piece

into

$$u(x) = \int_0^1 G(x,y)f(y)dy = \int_0^x y(1-x)f(y)dy + \int_x^1 x(1-y)f(y)dy$$

and apply Leibniz integral rule to each piece.

Proposition 2.1. Let $f : [0,1] \to \mathbb{R}$ be continuous on $[0,1]$, then u as defined in (2) is twice continuously differentiable and solves (1)

verifying solutions: check BCs

Proof. We split the proof into two steps.
Step 1: boundary condition: Check that $u(0) = u(1) = 0$ in the next homework.
Step 2: differential equation:
To verify $-u''(x) = f(x)$ for $0 < x < 1$, write

split integral and verify with leibniz

$$u(x) = \int_0^1 G(x,y)f(y)dy = \int_0^x y(1-x)f(y)dy + \int_x^1 x(1-y)f(y)dy$$

and apply Leibniz integral rule. We then compute

verifying solutions, use leibniz!!!

$$\frac{d}{dx}\int_0^x y(1-x)f(y)dy = x(1-x)f(x) + \int_0^x -yf(y)dy,$$
$$\frac{d}{dx}\int_x^1 x(1-y)f(y)dy = -x(1-x)f(x) + \int_x^1 (1-y)f(y)dy.$$

Adding the above terms together shows

apply FTOC here

$$u'(x) = \int_0^x -yf(y)dy + \int_x^1 (1-y)f(y)dy$$

The rest of the computation will be part of the homework. □

process:

1. show that BCs hold. plug into G

2. verify $-u'' = f$ by splitting into two integrals and then applying leibniz rule

maximum principle : let u be C^2. if $-u''(x) > 0$ for $x \in (0,1)$ and $u(0), u(1) \geq 0$ then $u(x) \geq 0$ for all $x \in [0,1]$

3. Uniqueness of solution: maximum principle

This discussion is parallel to Chapter 2.1.3 in the textbook, but will provide alternative proofs. The proofs in Chapter 2.1.3 of the textbook use the Green's function.

We'll now prove what is known as the weak maximum principle, which is the following statement.

Proposition 3.1 (maximum principle)**.** Let u be a twice continuously differentiable function such that

$$-u''(x) > 0 \text{ for } x \in (0,1),$$
$$u(0), u(1) \geq 0$$

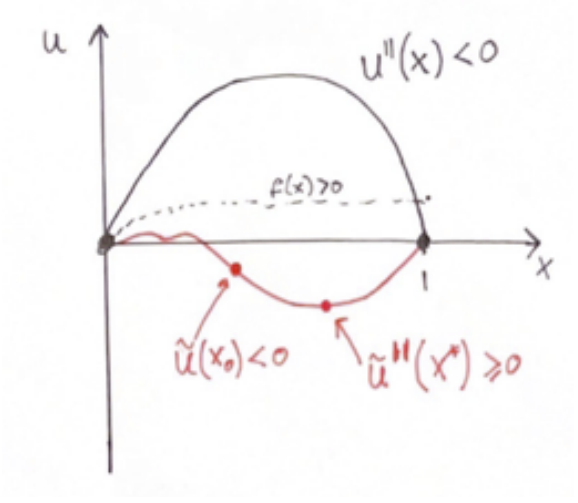
Then $u(x) \geq 0$ for all $x \in [0,1]$.

$-u''(x) > 0$ (concave)
$u(0), u(1) \geq 0$
$\Downarrow$
$u(x) \geq 0$ for all x

$u''(x) < 0$

ie. if u is concave and $u(0), u(1)$ are nonnegative, then $u(x)$ is also nonnegative

Picture of maximum principle: Since $-u''(x) > 0$, we have that $u''(x) < 0$ for all x, and u must be concave down on $[0,1]$. The picture of the function is below in black. We now suppose that a red function $\tilde{u}$ solves the above problem and has $\tilde{u}(x_0) < 0$. Since $\tilde{u}(0), \tilde{u}(1) \geq 0$, then $\tilde{u}$ must turn upwards at some point and thus be concave up, which would contradict $\tilde{u}''(x) < 0$.

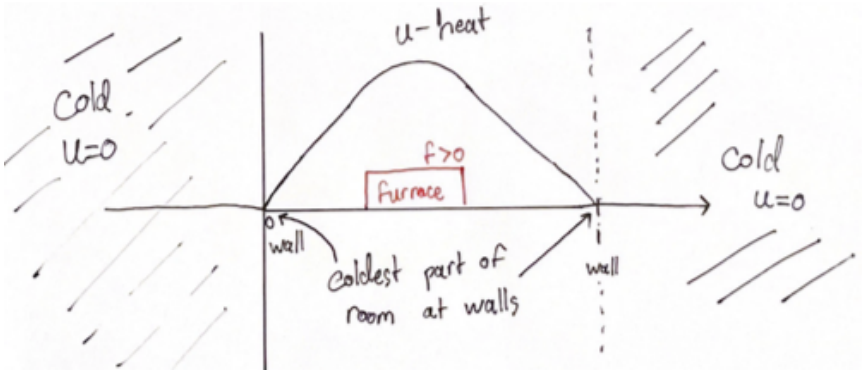


Physical intuition of maximum principle: Poisson's equation can also describe the distribution of heat in a room. If u stands for heat, then $u(0), u(1)$ are the outside temperature. If $-u''(x) = f(x)$, the f stands for the heat source, like a furnace. If $f(x) > 0$ everywhere, then we are heating the room from the inside. Maximum principle states that the coldest part of the room must be the walls.

WTS $u(x) \geq 0$

process : proving maximum principle
* proof by contradiction
AFSOC x_0 st. $u(x_0) < 0$ → implies there exists minimum
if not interior, show cannot be on boundaries

then apply calc II min/max arguments
this will contradict the pde
and we conclude proof



Proof. (By contradiction). The proof follows the picture. Suppose there is an x_0 such that $u(x_0) < 0$. Then, there must be an interior minimum. That is there must be an x^* such that $u(x^*) \leq u(y)$ for all y. Consequently, $u''(x^*) \geq 0$ and $-u''(x^*) \leq 0$. However, $-u''(x^*) > 0$, hence we have a contradiction and there cannot be an x_0 such that $u(x_0) < 0$. □

A consequence of the maximum principle is that the problem is stable as in the size of the solution u can be upper bounded by the size of $|f|$.

Corollary 3.1 (stability)**.** Suppose f is continuous on $[0,1]$. Let u solve (1). Then u satisfies

$$\max_{x\in[0,1]} |u(x)| \leq \max_{x\in[0,1]} |f(x)|$$

Proof. Let $M = \max_{x\in[0,1]} |f(x)|$. We break the proof into showing an upper bound $u(x) \leq M$ and lower bound $-M \leq u(x)$.

Step 1 (upper bound on u): Let $w(x) = Mx(1-x)$. The function w satisfies

$$0 \leq w(x) \leq M \text{ and } w''(x) = -2M$$

We define $\tilde{u}(x) = w(x) - u(x)$. Note that $\tilde{u}(x)$ solves

$$-\tilde{u}''(x) = -w''(x) - f(x) = 2M - f(x) > 0 \text{ for } x \in (0,1)$$
$$\tilde{u}(0) = w(0) = 0 \text{ and } \tilde{u}(1) = w(1) = 0.$$

Applying maximum principle to $\tilde{u}$ means $\tilde{u}(x) \geq 0$ and

$$u(x) \leq w(x) \leq M$$

for all $x \in [0,1]$.

Step 2 (lower bound on u): To show $-M \leq u(x)$ for all $x \in [0,1]$, define

$$\hat{u}(x) = u(x) + w(x)$$

and repeat the arguments of Step 1 for $\hat{u}$. Note that $\hat{u}(x)$ solves

$$-\hat{u}''(x) = -u''(x) - w''(x) = f(x) + 2M > 0 \text{ for } x \in (0,1)$$
$$\hat{u}(0) = w(0) = 0 \text{ and } \hat{u}(1) = w(1) = 0.$$

We can then apply maximum principle to $\hat{u}$ to get for all $x \in [0,1]$:

$$\hat{u}(x) \geq 0$$

and

$$u(x) \geq -w(x) \geq -M,$$

which is the desired lower bound. □

Corollary 3.2 (uniqueness of solutions)**.** Let u_1, u_2 be twice continuously differentiable solutions to (1). Then $u_1(x) = u_2(x)$ for all $x \in [0,1]$.

Proof. Let $v = u_1 - u_2$. Then v solves

$$-v''(x) = 0 \text{ for } x \in (0,1),$$
$$v(0) = v(1) = 0$$

Applying the stability result to v shows

$$\max_{x \in [0,1]} |v(x)| \leq 0$$

Hence, $v(x) = 0$ for all $x \in [0,1]$ and $u_1 = u_2$. □

Remark 3.1. There are two important points the proof of stability and uniqueness.

- If a linear differential equation only has terms containing u', u'', then adding constant like C to u will not change the fact that u solves the differential equation. Though $u + C$ may not satisfy the Dirichlet or Robin BC.
- For differential equations, we have the following general pattern:

Linear differential equation + stability = unique solutions

COMPUTATIONAL PDE LECTURE 4

LUCAS BOUCK

1. Outline of today

- Finish discussion of Maximum principle
- Finite difference solution of Poisson's equation in 1D

2. Uniqueness of solution: maximum principle

This discussion is parallel to Chapter 2.1.3 in the textbook, but will provide alternative proofs. The proofs in Chapter 2.1.3 of the textbook use the Green's function.

Recall we have been studying Poisson's equation 1 dimension:

$$\begin{cases} -u''(x) = f(x) & \text{for } x \in (0,1) \\ u(0) = u(1) = 0 \end{cases}, \tag{1}$$

and proved

Proposition 2.1 (maximum principle)**.** Let u be a twice continuously differentiable function such that

$$-u''(x) > 0 \text{ for } x \in (0,1),$$
$$u(0), u(1) \geq 0$$

Then $u(x) \geq 0$ for all $x \in [0,1]$.

A consequence of the maximum principle is that the problem is stable as in the size of the solution u can be upper bounded by the size of $|f|$.

Corollary 2.1 (stability)**.** Suppose f is continuous on $[0,1]$. Let u solve (1). Then u satisfies

$$\max_{x\in[0,1]} |u(x)| \leq \max_{x\in[0,1]} |f(x)|$$

Proof. Let $M = \max_{x\in[0,1]} |f(x)|$. We break the proof into showing an upper bound $u(x) \leq M$ and lower bound $-M \leq u(x)$.

Step 1 (upper bound on u): Let $w(x) = Mx(1-x)$. The function w satisfies

$$0 \leq w(x) \leq M \text{ and } w''(x) = -2M$$

w(x) is defined with these two properties in mind

needed for the following:

$$-\tilde{u}''(x) = -w''(x) - f(x)$$
$$= 2M - f(x) > 0$$

we know this is positive

this eventually allows us to make claims regarding $u(x) \leq M$ and $u(x) \geq -M$

Date: September 6, 2023.

for all $x \in [0,1]$. We define $\tilde{u}(x) = w(x) - u(x)$. Note that $\tilde{u}(x)$ solves

$$-\tilde{u}''(x) = -w''(x) - f(x) = 2M - f(x) > 0 \text{ for } x \in (0,1)$$
$$\tilde{u}(0) = w(0) = 0 \text{ and } \tilde{u}(1) = w(1) = 0.$$

Applying maximum principle to $\tilde{u}$ means $\tilde{u}(x) \geq 0$ and

$$u(x) \leq w(x) \leq M$$

for all $x \in [0,1]$.

Step 2 (lower bound on u): To show $-M \leq u(x)$ for all $x \in [0,1]$, define

$$\hat{u}(x) = u(x) + w(x)$$

and repeat the arguments of Step 1 for $\hat{u}$. □

Corollary 2.2 (uniqueness of solutions)**.** Let u_1, u_2 be twice continuously differentiable solutions to (1). Then $u_1(x) = u_2(x)$ for all $x \in [0,1]$.

Proof. Let $v = u_1 - u_2$. Then v solves

$$-v''(x) = 0 \text{ for } x \in (0,1),$$
$$v(0) = v(1) = 0$$

Apply the stability result to v to show $v \equiv 0$. □

Remark 2.1. There are two important points the proof of stability and uniqueness.

- If a linear differential equation only has terms containing u', u'', then adding constant like C to u will not change the fact that u solves the differential equation. Though $u + C$ may not satisfy the Dirichlet or Robin BC.
- For differential equations, we have the following general pattern:

$$\text{Linear differential equation} \ + \ \text{stability} \ = \ \text{unique solutions}$$

3. Finite Differences

Time permitting, we'll will start the discussion of finite differences.

Given a function f, how do we approximate it's derivative? From calculus, we remember the limit definition of derivative:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h},$$

which states that the tangent slope is the limit of the secant slopes.

The idea behind a finite difference is that although we may not know f', we can approximate it by fixing $h > 0$ rather than taking a limit:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}.$$

The above formula is what we call a **forward finite difference**. We could also look at

$$f'(x) \approx \frac{f(x) - f(x-h)}{h},$$

which would be a **backward finite difference**. Averaging these two approximations leads a **centered finite difference**:

$$f'(x) \approx \frac{1}{2}\left(\frac{f(x+h) - f(x)}{h} + \frac{f(x) - f(x-h)}{h}\right) = \frac{f(x+h) - f(x-h)}{2h}.$$

How might we construct a finite difference approximation for $f''(x)$? First, let's consider applying a forward finite difference to $f'(x)$:

$$f''(x) \approx \frac{f'(x+h) - f'(x)}{h}.$$

We don't know $f'(x)$ or $f'(x+h)$ but further use the approximations:

$$f'(x) \approx \frac{f(x) - f(x-h)}{h} \quad \text{(backward difference)}$$
$$f'(x+h) \approx \frac{f(x+h) - f(x)}{h} \quad \text{(backward difference)}$$

Subtracting these two approximations leads to

$$f'(x+h) - f'(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h}.$$

Further dividing by h gives us a potential approximation for the second derivative:

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} =: D_h^2 f(x). \tag{2}$$

The above approximation (2) will be our main tool for solving (1) numerically.

3.1. **Finite difference approximation of Poisson's equation.** We now apply (2) to approximate (1). We first start with a

- **mesh** or grid of $N+1$ evenly spaced points on the interval $[0,1]$ as $x_j = \frac{j}{N}$ for $i = 0, \ldots, N$.
- **mesh size** as $h = \frac{1}{N}$.
- $U : \{x_j\}_{j=0}^{N+1} \to \mathbb{R}$ is an unknown **grid function** that will approximate u.

At a point x_j for $0 < j < N+1$, we replace $u''(x_j)$ with $D_h^2 U(x_j)$. As a result, we replace

$$-u''(x_j) = f(x_j)$$

with

$$-D_h^2 U(x_j) = -\frac{U(x_{j+1}) - 2U(x_j) + U(x_{j-1})}{h^2} = f(x_j)$$

At x_0, x_{N+1}, we replace the boundary condition

$$u(0), u(1) = 0$$

with

$$U(x_0), U(x_N) = 0.$$

Combining the above two conditions leads to the following linear system of equations:

$$\underbrace{\begin{pmatrix} 1 & 0 & \dots & & \\ -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} & 0 & \dots \\ & \ddots & \ddots & \ddots & \\ \dots & 0 & -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} \\ & & \dots & 0 & 1 \end{pmatrix}}_{=:\mathbf{A}} \underbrace{\begin{pmatrix} U(x_0) \\ U(x_1) \\ \vdots \\ U(x_{N-1}) \\ U(x_N) \end{pmatrix}}_{=:\mathbf{U}} = \underbrace{\begin{pmatrix} 0 \\ f(x_1) \\ \vdots \\ f(x_{N-1}) \\ 0 \end{pmatrix}}_{\mathbf{f}}. \tag{3}$$

We can then solve $\mathbf{AU} = \mathbf{f}$ using a linear algebra solver like Gaussian elimination.

3.2. **Truncation error analysis: consistency.** By replacing $u''(x_j)$ with $D_h^2 U(x_j)$, we are making an approximation. A natural question to ask is how accurate is that approximation in terms of the number of grid points $N+1$ or the mesh size h. The first step to deriving an estimate for the error $|u(x_j) - U(x_j)|$ is to derive the **truncation error** τ^h, which is the error of plugging in the exact solution u into the discrete system (3):

$$\tau^h = \mathbf{f} - \mathbf{Au}$$

where $\mathbf{u}_j = u(x_j)$. Another way of writing the truncation error is $|\tau_j^h| = |u''(x_j) - D_h^2 u(x_j)|$. We say the discrete system (3) is **consistent** with (1) if the truncation error τ^h satisfies

$$\lim_{h \to 0} \|\tau^h\|_\infty = 0$$

where

$$\|\tau^h\|_\infty = \max_{0 \le j \le N+1} |\tau_j^h|$$

Proposition 3.1 (consistency and truncation error)**.** Suppose u solves is $C^4[0,1]$ and solves (1). Then for $1 \le j \le N$, τ_j^h satisfies

$$|\tau_j^h| \le \frac{h^2}{12} \max_{x \in [0,1]} |u^{(4)}(x)|$$

Proof. The proof relies heavily on Taylor expansion. For $u \in C^4[0,1]$, $x \in (0,1)$, and sufficiently small h, we have

$$u(x+h) = u(x) + hu'(x) + \frac{h^2}{2!}u''(x) + \frac{h^3}{3!}u^{(3)}(x) + \frac{h^4}{4!}u^{(4)}(h_1) \tag{4}$$

where $h_1 \in (x, x+h)$ and

$$u(x-h) = u(x) - hu'(x) + \frac{h^2}{2!}u''(x) - \frac{h^3}{3!}u^{(3)}(x) + \frac{h^4}{4!}u^{(4)}(h_2) \tag{5}$$

$h_2 \in (x-h, x)$. Adding (4) and (5) together yields

$$u(x+h) + u(x-h) = 2u(x) + h^2u''(x) + \frac{h^4}{4!}(u^{(4)}(h_1) + u^{(4)}(h_2)).$$

Hence,

$$\frac{u(x+h) - 2u(x) + u(x-h)}{h^2} = u''(x) + \frac{h^2}{4!}(u^{(4)}(h_1) + u^{(4)}(h_2)).$$

Subtracting $u''(x)$ from both sides and taking absolute value allows us to estimate:

$$\begin{aligned}\left|\frac{u(x+h) - 2u(x) + u(x-h)}{h^2} - u''(x)\right| &= \left|\frac{h^2}{4!}(u^{(4)}(h_1) + u^{(4)}(h_2))\right| \\ &\leq \frac{2h^2}{4!}\max_{z\in[0,1]}|u^{(4)}(z)| \\ &= \frac{h^2}{12}\max_{z\in[0,1]}|u^{(4)}(z)|\end{aligned}$$

The above estimate is true for any $x \in (0,1)$, so it is true for x_j. Hence

$$|\tau_j^h| \leq \frac{h^2}{12}\max_{z\in[0,1]}|u^{(4)}(z)|$$

for all j, which is the desired result. □

Definition 3.1 (rate of convergence)**.** A convergent sequence $z_N \to z$ as $N \to \infty$ converges with rate α if there is a $c > 0$ such that

$$|z_N - z| \leq c(1/N)^\alpha$$

Indexing with h, we say a sequence $z_h \to z$ as $h \to 0$ converges with rate α if there is a $c > 0$ such that

$$|z_h - z| \leq ch^\alpha.$$

Remark 3.1 (rate of convergence of truncation error)**.** We can see that $\tau_j^h \to 0$ with rate 2.

COMPUTATIONAL PDE LECTURE 5

LUCAS BOUCK

1. Outline of today

- Truncation error of our discretization of Poisson: consistency
- Convergence of the finite difference method assuming stability.

We have been studying

$$\begin{cases} -u''(x) = f(x) & \text{for } x \in (0,1) \\ u(0) = u_\ell, & \\ u(1) = u_r & \end{cases}, \tag{1}$$

for $u_\ell = u_r = 0$. The rest of the lecture will cover when u_ℓ, u_r are not necessarily 0.

2. Finite Differences Approximation of Poisson

Recall that our main tool for discretizing Poisson's equation was the finite difference approximation:

$$D_h^2 f(x) := \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}. \tag{2}$$

Using the above approximation (2), we discretized (1) with the equation at $x_j = hj$

$$-D_h^2 U^h(x_j) = -\frac{U^h(x_{j+1}) - 2U^h(x_j) + U^h(x_{j-1})}{h^2} = f(x_j),$$

and the equations

$$U^h(x_0) = u_\ell, U^h(x_N) = u_r$$

Date: September 8, 2023.

at x_0 and x_{N+1} respectively. Combining the above two conditions leads to the following linear system of equations:

$$\underbrace{\begin{pmatrix} 1 & 0 & \dots & & \\ -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} & 0 & \dots \\ & \ddots & \ddots & \ddots & \\ \dots & 0 & -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} \\ & & \dots & 0 & 1 \end{pmatrix}}_{=:\mathbf{A}^h} \underbrace{\begin{pmatrix} U^h(x_0) \\ U^h(x_1) \\ \vdots \\ U^h(x_{N-1}) \\ U^h(x_N) \end{pmatrix}}_{=:\mathbf{U}^h} = \underbrace{\begin{pmatrix} u_\ell \\ f(x_1) \\ \vdots \\ f(x_{N-1}) \\ u_r \end{pmatrix}}_{\mathbf{f}^h}. \tag{3}$$

We now begin to analysis this method.

2.1. **Truncation error analysis: consistency.** By replacing $u''(x_j)$ with $D_h^2 U^h(x_j)$, we are making an approximation. A natural question to ask is how accurate is that approximation in terms of the number of grid points $N+1$ or the mesh size h. The first step to deriving an estimate for the error $|u(x_j) - U^h(x_j)|$ is to derive the **truncation error** $\boldsymbol{\tau}^h$, which is the error of plugging in the exact solution u into the discrete system (3):

$$\boldsymbol{\tau}^h = \mathbf{f}^h - \mathbf{A}^h \mathbf{u}$$

where $\mathbf{u}_j = u(x_j)$. Another way of writing the truncation error is $|\boldsymbol{\tau}_j^h| = |u''(x_j) - D_h^2 u(x_j)|$. We say the discrete system (3) is **consistent** with (1) if the truncation error $\boldsymbol{\tau}^h$ satisfies

$$\lim_{h \to 0} \|\boldsymbol{\tau}^h\|_\infty = 0$$

where

$$\|\boldsymbol{\tau}^h\|_\infty = \max_{0 \le j \le N+1} |\boldsymbol{\tau}_j^h|$$

Proposition 2.1 (consistency and truncation error)**.** Suppose u solves is $C^4[0,1]$ and solves (1). Then for $1 \le j \le N-1$, $\boldsymbol{\tau}_j^h$ satisfies

$$|\boldsymbol{\tau}_j^h| \le \frac{h^2}{12} \max_{x \in [0,1]} |u^{(4)}(x)|$$

Proof. The proof relies heavily on Taylor expansion. For $u \in C^4[0,1]$, $x \in (0,1)$, and sufficiently small h, we have

$$u(x+h) = u(x) + hu'(x) + \frac{h^2}{2!}u''(x) + \frac{h^3}{3!}u^{(3)}(x) + \frac{h^4}{4!}u^{(4)}(h_1) \tag{4}$$

where $h_1 \in (x, x+h)$ and

$$u(x-h) = u(x) - hu'(x) + \frac{h^2}{2!}u''(x) - \frac{h^3}{3!}u^{(3)}(x) + \frac{h^4}{4!}u^{(4)}(h_2) \tag{5}$$

$h_2 \in (x-h, x)$. Adding (4) and (5) together yields

$$u(x+h) + u(x-h) = 2u(x) + h^2 u''(x) + \frac{h^4}{4!}(u^{(4)}(h_1) + u^{(4)}(h_2)).$$

Hence,

$$\frac{u(x+h) - 2u(x) + u(x-h)}{h^2} = u''(x) + \frac{h^2}{4!}(u^{(4)}(h_1) + u^{(4)}(h_2)).$$

Subtracting $u''(x)$ from both sides and taking absolute value allows us to estimate:

$$\begin{aligned}\left|\frac{u(x+h) - 2u(x) + u(x-h)}{h^2} - u''(x)\right| &= \left|\frac{h^2}{4!}(u^{(4)}(h_1) + u^{(4)}(h_2))\right| \\ &\leq \frac{2h^2}{4!}\max_{z\in[0,1]}|u^{(4)}(z)| = \frac{h^2}{12}\max_{z\in[0,1]}|u^{(4)}(z)|\end{aligned}$$

The above estimate is true for any $x \in (0,1)$, so it is true for x_j. Hence

$$|\tau_j^h| \leq \frac{h^2}{12}\max_{z\in[0,1]}|u^{(4)}(z)|$$

for all j, which is the desired result. □

Definition 2.1 (rate of convergence)**.** A convergent sequence $z_N \to z$ as $N \to \infty$ converges with rate α if there is a $c > 0$ such that

$$|z_N - z| \leq c(1/N)^\alpha$$

Indexing with h, we say a sequence $z_h \to z$ as $h \to 0$ converges with rate α if there is a $c > 0$ such that

$$|z_h - z| \leq ch^\alpha.$$

Remark 2.1 (rate of convergence of truncation error)**.** We can see that $\tau_j^h \to 0$ with rate 2.

Remark 2.2 (big O notation)**.** We say that $z_h = O(a_h)$ as $h \to 0$ if there is a $h_0, c > 0$ such that

$$z_h \leq ca_h$$

for all $h < h_0$. For example, $\tau_j^h = O(h^2)$.

big O : $z_h = O(a_h)$ as $h \to 0$ if there is a $h_0, c > 0$ such that $z_h \leq ca_h$ for all $h < h_0$

convergence :
a sequence $z_h \to z$ as $h \to 0$ converges with rate α if there is a $c > 0$ such that
$|z_h - z| \leq ch^\alpha$

$\tau_j^h = O(h^2)$

because $|\tau_j^h| \leq \frac{1}{12}h^2 \max|u^{(4)}(z)|$
we see τ_j^h converges with rate $\alpha = 2$

2.2. **Discrete maximum principle: stability.** The rest of the discussion will assume $u_r = u_\ell = 0$ though these results can be adapted if $u_r \neq 0$ and $u_\ell \neq 0$, which will be an assignment problem for the continuous problem. The book also presents different proofs based on discrete Green's function.

So far, we have shown that the exact solution u solves the discrete equation with an additional truncation error term that is $O(h^2)$. However, we have so far not said anything about the actual error:

$$\max_{0\leq j\leq N} |u(x_j) - U^h(x_j)|.$$

What we'll need is a stability result, which we'll prove in Monday's lecture.

Proposition 2.2 (stability of the finite difference scheme)**.** For $u_\ell = u_r = 0$, the discrete scheme is stable in the sense that

$$\max_{0\leq j\leq N} |U^h(x_j)| \leq \max_{1\leq j\leq N} |f(x_j)|$$

In other words, if $\mathbf{A}^h\mathbf{U}^h = \mathbf{f}^h$, then

$$\|\mathbf{U}^h\|_\infty \leq \|\mathbf{f}^h\|_\infty$$

Additionally, the right hand side of the above inequality is independent of h.

We will prove this later. There are two important corollaries of stability. The first is that there exist unique discrete solutions U^h.

Corollary 2.1 (existence and uniqueness of discrete solutions)**.** For $u_\ell = u_r = 0$, there exists a unique $\mathbf{U}^h$ such that solves the system of linear equations (3)

Proof. Since $\mathbf{A}^h$ is a square matrix, existence of solutions to (3) is equivalent to uniqueness of solutions by the Fundamental Theorem of Linear Algebra. To show uniqueness let $\mathbf{U}^h, \mathbf{V}^h$ solve

$$\mathbf{A}^h\mathbf{U}^h = \mathbf{f}^h, \quad \mathbf{A}^h\mathbf{V}^h = \mathbf{f}^h.$$

Subtracting these two equations yields

$$\mathbf{A}^h(\mathbf{U}^h - \mathbf{V}^h) = \mathbf{0}.$$

Applying the stability result shows that

$$\|\mathbf{U}^h - \mathbf{V}^h\|_\infty \leq \|\mathbf{0}\|_\infty$$

and $\mathbf{U}^h = \mathbf{V}^h$. □

The second important corollaries is the desired error estimate:

Theorem 2.1 (convergence and error estimate)**.** Let $u_\ell = u_r = 0$. Let $u \in C^4[0,1]$ be a solution of (1) and let U^h be a solution of (3), then we have

$$\max_{0\le j\le N} |u(x_j) - U^h(x_j)| \le \frac{h^2}{12} \max_{x\in[0,1]} |u^{(4)}(x)|$$

Proof. Let $\mathbf{U}^h_j = U^h(x_j)$, which is the solution to

$$\mathbf{A}^h\mathbf{U}^h = \mathbf{f}^h.$$

Let $\mathbf{u}_j = u(x_j)$ be the vector of the exact solution u evaluated at x_j. We saw that $\mathbf{u}$ solves

$$\mathbf{A}^h\mathbf{u} = \mathbf{f}^h + \boldsymbol{\tau}^h,$$

where $\boldsymbol{\tau}^h$ is the truncation error and

$$\|\boldsymbol{\tau}^h\|_\infty \le \frac{h^2}{12} \max_{z\in[0,1]} |u^{(4)}(z)|.$$

Consider the error vector $\mathbf{e}^h = \mathbf{u} - \mathbf{U}^h$. By subtracting the equations for $\mathbf{u}$ and $\mathbf{U}^h$, we see that

$$\mathbf{A}^h\mathbf{e}^h = \boldsymbol{\tau}^h.$$

We apply the stability result to the above problem to see that

$$\max_{0\le j\le N} |u(x_j) - U^h(x_j)| = \|\mathbf{e}^h\|_\infty \le \|\boldsymbol{\tau}^h\|_\infty.$$

Recall the truncation error satisfies:

$$\|\boldsymbol{\tau}^h\|_\infty \le \frac{h^2}{12} \max_{x\in[0,1]} |u^{(4)}(x)|,$$

so

$$\max_{0\le j\le N} |u(x_j) - U^h(x_j)| \le \frac{h^2}{12} \max_{x\in[0,1]} |u^{(4)}(x)|.$$

which is the desired error estimate. □

Remark 2.3 (Lax postulate)**.** The lecture today shows an important principle of numerical methods for linear differential equations:

$$\text{Consistency} + \text{Stability} \implies \text{Convergence}$$

Another name for this is the **Lax equivalence theorem**.

COMPUTATIONAL PDE LECTURE 6

LUCAS BOUCK

1. Outline of today

- Discrete maximum principle and stability
- Handling Neumann boundary conditions

We have been studying

$$\begin{cases} -u''(x) = f(x) & \text{for } x \in (0,1) \\ u(0) = u_\ell, \\ u(1) = u_r \end{cases}, \tag{1}$$

for $u_\ell = u_r = 0$.

2. Finite Differences Approximation of Poisson

Recall that our main tool for discretizing Poisson's equation was the finite difference approximation:

$$D_h^2 f(x) := \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}. \tag{2}$$

Using the above approximation (2), we discretized (1) with the equation at $x_j = hj$

$$-D_h^2 U^h(x_j) = -\frac{U^h(x_{j+1}) - 2U^h(x_j) + U^h(x_{j-1})}{h^2} = f(x_j),$$

and the equations

$$U^h(x_0) = u_\ell, U^h(x_N) = u_r$$

Date: September 11, 2023.

at x_0 and x_{N+1} respectively. Combining the above two conditions leads to the following linear system of equations:

$$\underbrace{\begin{pmatrix} 1 & 0 & \dots & & \\ -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} & 0 & \dots \\ & \ddots & \ddots & \ddots & \\ \dots & 0 & -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} \\ & & \dots & 0 & 1 \end{pmatrix}}_{=:\mathbf{A}^h} \underbrace{\begin{pmatrix} U^h(x_0) \\ U^h(x_1) \\ \vdots \\ U^h(x_{N-1}) \\ U^h(x_N) \end{pmatrix}}_{=:\mathbf{U}^h} = \underbrace{\begin{pmatrix} u_\ell \\ f(x_1) \\ \vdots \\ f(x_{N-1}) \\ u_r \end{pmatrix}}_{\mathbf{f}^h}. \tag{3}$$

What we'll need is a stability result, which we'll prove in Monday's lecture.

Proposition 2.1 (stability of the finite difference scheme)**.** For $u_\ell = u_r = 0$, the discrete scheme is stable in the sense that

$$\max_{0\le j\le N} |U^h(x_j)| \le \max_{1\le j\le N} |f(x_j)|$$

In other words, if $\mathbf{A}^h\mathbf{U}^h = \mathbf{f}^h$, then

$$\|\mathbf{U}^h\|_\infty \le \|\mathbf{f}^h\|_\infty$$

Additionally, the right hand side of the above inequality is independent of h.

The important consequence of stability was the error estimate:

Theorem 2.1 (convergence and error estimate)**.** Let $u_\ell = u_r = 0$. Let $u \in C^4[0,1]$ be a solution of (1) and let U^h be a solution of (3), then we have

$$\max_{0\le j\le N} |u(x_j) - U^h(x_j)| \le \frac{h^2}{12} \max_{x\in[0,1]} |u^{(4)}(x)|$$

2.0.1. *Discrete maximum principle and proof of stability.*

Lemma 2.1 (discrete maximum principle)**.** Let U^h be the discrete solution to (3), and assume $u_r, u_\ell \ge 0$ and $f(x) > 0$ for all $x \in [0,1]$. Then, $U^h(x_j) \ge 0$ for all $j = 0, \dots, N$.

Proof. The proof follows similarly to that of the continuous problem. We proceed by contradiction. Suppose there is an $j_0 \in \{1, \dots, N-1\}$ such that $U^h(x_{j_0}) < 0$. There then exists a $J \in \{1, \dots, N-1\}$ such that $U^h(x_J) < 0$ and $U^h(x_J) \le U^h(x_j)$ for all $j \in \{0, \dots, N\}$. We now look compute $-D_h^2 U^h(x_J)$:

$$-D_h^2 U^h(x_J) = \frac{-U^h(x_{J+1}) + 2U^h(x_J) - U^h(x_{J-1})}{h^2}.$$

Since $U^h(x_J) \le U^h(x_{J\pm1})$, we have

$$\frac{-U^h(x_{J+1}) + 2U^h(x_J) - U^h(x_{J-1})}{h^2} \le \frac{-U^h(x_{J+1}) + U^h(x_{J+1}) + U^h(x_{J-1}) - U^h(x_{J-1})}{h^2} = 0.$$

Hence, $-D_h^2 U^h(x_J) \leq 0$. By assumption, $-D_h^2 U^h(x_J) = f(x_J) > 0$. Thus, $0 < -D_h^2 U^h(x_J) \leq 0$, which is a contradiction. □

In fact, the discrete maximum principle is much more general than just Poisson's equation which is part of the homework.

We can prove the desired stability result.

Proof of discrete stability. The proof also follows similarly to the proof of stability of the continuous problem. We first define $M = \max_{x\in[0,1]} |f(x)|$.

Let $\tilde{U}^h : \{x_j\}_{j=0}^N$ be defined by

$$W^h(x_j) = (M+\varepsilon)x_j(1-x_j)$$

for $\varepsilon > 0$. One can check that for $0 \leq x_j \leq 1$:

$$D_h^2 W^h(x_j) = -(2M+2\varepsilon) \text{ and } 0 \leq W^h(x_j) \leq M+\varepsilon.$$

We repeat the arguments of the stability result for the continuous Poisson equation.

Step 1. (upper bound) We define $\tilde{U}^h = W^h - U^h$. Then,

$$-D_h^2 \tilde{U}^h(x_j) = -D_h^2 W^h(x_j) + D_h^2 U^h(x_j) = (2M+2\varepsilon) - f(x_j) \geq M+\varepsilon > 0,$$

and

$$\tilde{U}^h(0) = W^h(0) = 0.$$

Applying the discrete maximum principle yields

$$0 \leq \tilde{U}^h(x_j) \implies U^h(x_j) \leq W^h(x_j) \leq M+\varepsilon.$$

Step 2. (lower bound) We define $\hat{U}^h = W^h + U^h$. Then, apply the arguments of Step 1 to get that $U^h(x_j) \geq -M-\varepsilon$.

Note that we have shown $|U^h(x_j)| \leq M+\varepsilon$. Since $\varepsilon > 0$ was arbitrary, we can take a limit $\varepsilon \to 0$ to get $|U^h(x_j)| \leq M$, which is the desired stability result. □

COMPUTATIONAL PDE LECTURE 7

LUCAS BOUCK

1. Outline of today

- Discrete maximum principle and stability
- Handling Neumann boundary conditions

2. Neumann Boundary Conditions

How would we implement a Neumann BC for the following problem?

$$\begin{cases} -u''(x) & = f(x) \\ u'(0) = g_\ell \\ u(1) = u_r \end{cases} \tag{1}$$

how would we handle neumann condition?

One option would be to have the discretized equation use a forward difference:

$$\frac{U^h(x_1) - U^h(x_0)}{h} = g_\ell$$

The issue with this approximation is that the truncation error is for some $C > 0$:

$$\left| \frac{u(x_1) - u(x_0)}{h} \right| = Ch \max_{z \in [0,1]} |u''(z)|,$$

using forward diff for the BC would have reduced convergence rate

which means we'd potentially have a reduced convergence rate.

2.1. Trick: Ghost point.

The neat trick is to use a technique called a ghost point, which utilizes a centered finite difference at 0:

$$\frac{U^h(x_1) - U^h(x_{-1})}{2h} = g_\ell$$

centered diff

ghost point x_{-1}

where $x_{-1} = -h$. This centered difference does not give us a useable equation, but we can use the differential equation approximation to write

$$\frac{-U^h(x_1) + 2U^h(x_0) - U^h(x_{-1})}{h^2} = f(x_0)$$

then use our approx and substitute for $U^h(x_{-1})$

Date: September 13, 2023.

ghost point : approx $u'(0) = g_\ell$ with the following

$$\frac{U^h(x_1) - U^h(x_{-1})}{h^2} = g_\ell$$

Since $U^h(x_{-1}) = U^h(x_1) - 2hg_\ell$, we have

$$\frac{2U^h(x_0) - 2U^h(x_1)}{h^2} = f(x_0) - \frac{2g_\ell}{h}$$

and the resulting linear system is:

$$\underbrace{\begin{pmatrix} \frac{2}{h^2} & \frac{-2}{h^2} & \dots & & \\ -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} & 0 & \dots \\ & \ddots & \ddots & \ddots & \\ \dots & 0 & -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} \\ & & \dots & 0 & 1 \end{pmatrix}}_{=:\mathbf{A}^h} \underbrace{\begin{pmatrix} U^h(x_0) \\ U^h(x_1) \\ \vdots \\ U^h(x_{N-1}) \\ U^h(x_N) \end{pmatrix}}_{=:\mathbf{U}^h} = \underbrace{\begin{pmatrix} f(x_0) - \frac{2g_\ell}{h} \\ f(x_1) \\ \vdots \\ f(x_{N-1}) \\ u_r \end{pmatrix}}_{\mathbf{f}^h}. \tag{2}$$

Remark 2.1. Note that the matrix $\mathbf{A}^h$ is no longer symmetric. This can be fixed by multiplying

$$\frac{2U^h(x_0) - 2U^h(x_1)}{h^2} = f(x_0) - \frac{2g_\ell}{h}$$

by $\frac{1}{2}$ to get

$$\frac{U^h(x_0) - U^h(x_1)}{h^2} = \frac{1}{2}f(x_0) - \frac{g_\ell}{h}$$

and the matrix system is:

$$\underbrace{\begin{pmatrix} \frac{1}{h^2} & \frac{-1}{h^2} & \dots & & \\ -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} & 0 & \dots \\ & \ddots & \ddots & \ddots & \\ \dots & 0 & -\frac{1}{h^2} & \frac{2}{h^2} & -\frac{1}{h^2} \\ & & \dots & 0 & 1 \end{pmatrix}}_{=:\mathbf{A}^h} \underbrace{\begin{pmatrix} U^h(x_0) \\ U^h(x_1) \\ \vdots \\ U^h(x_{N-1}) \\ U^h(x_N) \end{pmatrix}}_{=:\mathbf{U}^h} = \underbrace{\begin{pmatrix} \frac{1}{2}f(x_0) - \frac{g_\ell}{h} \\ f(x_1) \\ \vdots \\ f(x_{N-1}) \\ u_r \end{pmatrix}}_{\mathbf{f}^h} \tag{3}$$

which is now symmetric.

2.2. **Another approach.** Another approach would be to discretize the Neumann BC with the following finite difference:

$$u'(x_0) \approx \frac{-3u(x_0) + 4u(x_1) - u(x_2)}{2h},$$

which you'll see in the HW satisfies:

$$u'(x_0) = \frac{-3u(x_0) + 4u(x_1) - u(x_2)}{2h} + O(h^2).$$

One drawback of this method is that you cannot make the resulting matrix $\mathbf{A}^h$ symmetric.

3. Variational Principle for Poisson's Equation

We now conclude with a variational principle for Poisson equation. Consider that the discrete $\mathbf{U}^h$ solves the following system of linear equations.

$$\mathbf{A}^h\mathbf{U}^h = \mathbf{f}^h$$

Observe that if

$$\mathbf{U}^h \text{ minimizes } \mathbf{v}^h \mapsto \frac{1}{2}\mathbf{v}^{hT}\mathbf{A}^h\mathbf{v}^h - \mathbf{f}^h\mathbf{v}^h$$

then

$$\mathbf{A}^h\mathbf{U}^h = \mathbf{f}^h.$$

In fact we have an analogous result for $\mathbf{u}$ solving Poisson's equation.

Proposition 3.1. Suppose $u \in C^1[0,1]$ subject to $u(0) = u(1) = 0$ minimizes

$$E[w] = \int_\Omega \frac{1}{2}|w'(x)|^2 - f(x)w(x)dx$$

over all possible $w \in C^1[0,1]$ subject to $w(0) = w(1) = 0$, then u satisfies

$$\int_\Omega u'(x)v'(x)dx = \int_\Omega f(x)v(x)dx$$

for all $v \in C^1[0,1]$ subject to $v(0) = v(1) = 0$. As shown in the HW, this is the weak form of Poisson's equation:

$$\begin{cases} -u''(x) = f(x) x \in (0,1) \\ u(0) = u(1) = 0 \end{cases}$$

Proof. Let u minimize E and let $w^t = u + tv$, where $t > 0$ and $v \in C^1[0,1]$ subject to $v(0) = v(1) = 0$. Then $w^t \in C^1[0,1]$ and $w^t(0) = w^t(1) = 0$. Since u is a minimizer we have,

$$0 \le E[w^t] - E[u] = E[u + tv] - E[u]$$

We now expand $E[u + tv]$:

$$\begin{aligned} E[u + tv] &= \int_\Omega \frac{1}{2}|u'(x) + tv'(x)|^2 - f(x)(u(x) + tv(x))dx \\ &= \int_\Omega \frac{1}{2}|u'(x)|^2 + \frac{t^2}{2}|v'(x)|^2 + tu'(x)v'(x) - f(x)(u(x) + tv(x))dx \end{aligned}$$

Subtracting $E[u]$ yields and dividing by t yields

$$\frac{E[u + tv] - E[u]}{t} = \int_\Omega \frac{t}{2}|v'(x)|^2 + u'(x)v'(x) - f(x)v(x)dx$$

Taking a limit as $t \to 0$ leads to

$$\int_\Omega u'(x)v'(x) - f(x)v(x)dx \geq 0$$

Repeating the argument for $w^t = u - tv$ also shows

$$\int_\Omega u'(x)v'(x) - f(x)v(x)dx \leq 0$$

Hence

$$\int_\Omega u'(x)v'(x) - f(x)v(x)dx = 0,$$

which is the weak form. □

The reverse direction is also true.

Proposition 3.2. Suppose $u \in C^1[0,1]$ satisfies

$$\int_\Omega u'(x)v'(x)dx = \int_\Omega f(x)v(x)dx$$

for all $v \in C^1[0,1]$ subject to $v(0) = v(1) = 0$. Then, u minimizes

$$E[w] = \int_\Omega \frac{1}{2}|w'(x)|^2 - f(x)w(x)dx$$

over all possible $w \in C^1[0,1]$ subject to $w(0) = w(1) = 0$.

Proof. Let $w \in C^1[0,1]$ subject to $w(0) = w(1) = 0$. Let u solve the weak form of Poisson's equation. We let $v = w - u$, which satisfies $v(0) = v(1) = 0$. Then,

$$E[w] - E[u] = E[u+v] - E[u]$$

$$= \int_\Omega \frac{1}{2}|u'(x)|^2 + \frac{1}{2}|v'(x)|^2 + u'(x)v'(x) - f(x)(u(x)+v(x))dx - \int_\Omega \frac{1}{2}|u'(x)|^2 - f(x)u(x)dx$$

$$= \int_\Omega \frac{1}{2}|v'(x)|^2 dx + \underbrace{\int_\Omega u'(x)v'(x) - f(x)v(x)dx}_{=0 \text{ by assumption}}$$

$$= \int_\Omega \frac{1}{2}|v'(x)|^2 dx \geq 0$$

Hence, $E[w] \geq E[u]$ for all $w \in C^1[0,1]$ subject to $w(0) = w(1) = 0$. □

These variational principles work for much more general problems. For instance, minimizing the energy:

$$E[u] = \int_\Omega \frac{1}{2}|u'(x)|^2 + f(u(x))dx$$

leads to the following weak form:

$$\int_\Omega u'(x)v'(x) + f'(u(x))v(x)dx = 0,$$

and whose strong form is

$$-u''(x) + f'(u(x)) = 0.$$

In HW2, the PDE

$$-u''(x) + \frac{1}{\varepsilon^2}u(x)(u(x)^2 - 1) = 0$$

is a PDE corresponding to the energy:

$$\int_\Omega \frac{1}{2}|u'(x)|^2 + \frac{1}{\varepsilon^2}(u(x)^2 - 1)^2 dx$$

COMPUTATIONAL PDE LECTURE 8

LUCAS BOUCK

1. Outline of today

- Derive the heat equation
- Derive energy estimates of the heat equation

2. Derive the heat equation

We'll start with deriving the heat equation on the real line. That is the internal heat energy density $u : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ solves

$$u_t(t,x) - k u_{xx}(t,x) = f(t,x). \tag{1}$$

u: heat density, f: external source

where f is a heat source or sink. We first start with a point x_0 and an interval $(x_0 - \delta, x_0 + \delta)$. We first write the **conservation of energy**

conservation of energy

(total change in energy) = (total energy change from external sources) − (energy leaving)

which can be expressed mathematically as

$$\frac{d}{dt}\int_{x_0-\delta}^{x_0+\delta} u(t,x)dx = \int_{x_0-\delta}^{x_0+\delta} f(t,x)dx - j(t,x_0+\delta) + j(t,x_0-\delta).$$

energy leaving

Here, j denotes the **heat flux**.

We do cannot determine a precise form of the heat flux a priori, but we assume the heat flux satisfies an empirical law. Specifically, we assume **Fourier's law of heat**, which states there is a constant $k > 0$ such that

$$j(t,x) = -k u_x(t,x).$$

fourier's law of heat : empirically observed

Intuitively, this means heat flows from hot areas to cold areas. The conservation of energy now reads:

$$\frac{d}{dt}\int_{x_0-\delta}^{x_0+\delta} u(t,x)dx = \int_{x_0-\delta}^{x_0+\delta} f(t,x)dx + k u_x(t,x_0+\delta) - k u_x(t,x_0-\delta).$$

Date: September 15, 2023.

We now use fundamental theorem of calculus (this would be divergence theorem in higher dimensions), to write

$$ku_x(t, x_0 + \delta) - ku_x(t, x_0 - \delta) = \int_{x_0-\delta}^{x_0+\delta} ku_{xx}(t, x)dx,$$

so

$$\frac{d}{dt}\int_{x_0-\delta}^{x_0+\delta} u(t, x)dx = \int_{x_0-\delta}^{x_0+\delta} f(t, x)dx + \int_{x_0-\delta}^{x_0+\delta} ku_{xx}(t, x)dx.$$

We also apply Leibniz integral rule to the left hand side to get

$$\frac{d}{dt}\int_{x_0-\delta}^{x_0+\delta} u(t, x)dx = \int_{x_0-\delta}^{x_0+\delta} u_t(t, x)dx,$$

and

$$\int_{x_0-\delta}^{x_0+\delta} u_t(t, x)dx = \int_{x_0-\delta}^{x_0+\delta} f(t, x)dx + \int_{x_0-\delta}^{x_0+\delta} ku_{xx}(t, x)dx.$$

Rearranging and dividing both sides by 2δ leads to

$$\frac{1}{2\delta}\int_{x_0-\delta}^{x_0+\delta} u_t(t, x) - ku_{xx}(t, x)dx = \frac{1}{2\delta}\int_{x_0-\delta}^{x_0+\delta} f(t, x)dx.$$

Recall that for continuous functions:

$$\lim_{\delta\to 0}\frac{1}{2\delta}\int_{x_0-\delta}^{x_0+\delta} f(t, x)dx = f(t, x_0),$$

and we have the heat equation (1).

3. Energy estimates for the heat equation

Recall that for Poissons equation, we answered the following questions:

- Existence and construction of solutions: Green's functions
- Stability: maximum principle
- Uniqueness of solutions: maximum principle

Today, we'll address the stability as well as uniqueness of solutions to the heat equation by looking at what are called **energy estimates**.

Proposition 3.1 (energy estimate)**.** Let u be a C^2 solution to the heat equation with homogenous Dirichlet boundary conditions.

$$(2)\qquad \begin{cases} u_t(t, x) - u_{xx}(t, x) = f(t, x), & t > 0 \text{ and } x \in (0, 1) \\ u(0) = u(1) = 0, & \text{(boundary condition)} \\ u(0, x) = u_0(x) & \text{(initial condition)} \end{cases}$$

Then, for all $T > 0$, we have

$$\int_0^1 |u(T,x)|^2 dx + \int_0^T \int_0^1 |u_x(t,x)|^2 dx dt \leq \int_0^1 |u_0(x)|^2 dx + \int_0^T \int_0^1 |f(t,x)|^2 dx dt \tag{3}$$

Proof. We begin by multiplying the heat equation by $u(t,x)$ and integrating from $x = 0$ to $x = 1$:

$$\int_0^1 u_t(t,x)u(t,x)dx - \int_0^1 u_{xx}(t,x)u(t,x)dx = \int_0^1 f(t,x)u(t,x)dx$$

We now make a few observations about this equation:

- The first time can be written as a pure time derivative using the chain rule and Leibniz integral rule:

$$\int_0^1 u_t(t,x)u(t,x)dx = \frac{1}{2}\int_0^1 \frac{\partial}{\partial t}|u(t,x)|^2 dx = \frac{d}{dt}\frac{1}{2}\int_0^1 |u(t,x)|^2 dx$$

- Integrating the second term by parts leads to:

$$-\int_0^1 u_{xx}(t,x)u(t,x)dx = -u_x(t,1)u(t,1) + u_x(t,0)u(t,0) + \int_0^1 u_x(t,x)u_x(t,x)dx$$
$$= \int_0^1 |u_x(t,x)|^2 dx$$

  Thus, our equality is now

$$\frac{d}{dt}\frac{1}{2}\int_0^1 |u(t,x)|^2 dx + \int_0^1 |u_x(t,x)|^2 dx = \int_0^1 f(t,x)u(t,x)dx$$

- The term on the right hand side can be written in a way that can be controlled by terms on the left hand side. We first use a fact about real numbers, which is for any $a, b \in \mathbb{R}$, we have

$$(a-b)^2 \geq 0.$$

  Expanding the quadratic leads to

$$a^2 + b^2 - 2ab \geq 0.$$

  Rearranging this inequality yields what is sometimes called **Young's inequality**:

$$ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2.$$

  Hence,

$$\int_0^1 f(t,x)u(t,x)dx \leq \frac{1}{2}\int_0^1 |f(t,x)|^2 dx + \frac{1}{2}\int_0^1 |u(t,x)|^2 dx.$$

poincare's : $\int_0^1 |u(t,x)|^2 dx \le \int_0^1 |u_x(t,x)|^2 dx$

The other useful inequality is to control the size of u with its derivative u_x, which is known as **Poincare's inequality**:

$$\int_0^1 |u(t,x)|^2 dx \le \int_0^1 |u_x(t,x)|^2 dx,$$

ie. a function cannot grow faster than its derivative

which we'll prove later. Using Poincare leads us to

$$\int_0^1 f(t,x)u(t,x)dx \le \frac{1}{2}\int_0^1 f(t,x)^2 dx + \frac{1}{2}\int_0^1 |u_x(t,x)|^2 dx.$$

replaced u with u_x from poincare

Combing all the equalities and estimates yields:

$$\frac{d}{dt}\frac{1}{2}\int_0^1 |u(t,x)|^2 dx + \int_0^1 |u_x(t,x)|^2 dx \le \frac{1}{2}\int_0^1 f(t,x)^2 dx + \frac{1}{2}\int_0^1 |u_x(t,x)|^2 dx.$$

We now absorb the the integral of $(u_x)^2$ onto the left hand side, multiply everything by 2, and integrate in time from 0 to T:

$$\int_0^T \frac{d}{dt}\int_0^1 |u(t,x)|^2 dx dt + \int_0^T \int_0^1 |u_x(t,x)|^2 dx dt \le \int_0^T \int_0^1 f(t,x)^2 dx dt.$$

Realizing

$$\int_0^T \frac{d}{dt}\int_0^1 |u(t,x)|^2 dx dt = \int_0^1 |u(T,x)|^2 dx - \int_0^1 |u(0,x)|^2 dx$$

from FTOC, but our function inside has an integral

finishes the proof. □

A corollary of the energy estimate is that solutions to the heat equation are unique.

energy estimate used to show uniqueness

Corollary 3.1 (uniqueness of solutions)**.** C^2 solutions to (2) are unique.

Proof. Let u_1, u_2 be C^2 solutions to (2). Then the difference $v = u_1 - u_2$ solves

$$\begin{cases} v_t(t,x) - v_{xx}(t,x) = 0, & t > 0 \text{ and } x \in (0,1) \\ v(0) = v(1) = 0, \\ v(0,x) = 0 \end{cases}$$

from linearity

Applying the energy estimates to v shows $v = 0$. □

Lemma 3.1 (Poincare's inequality)**.** Let $f \in C^1[0,1]$ satisfy $f(0) = 0$, then

$$\int_0^1 |f(x)|^2 dx \le \int_0^1 |f'(x)|^2 dx$$

Proof. We can write

$$f(x) = \cancel{f(0)} + \int_0^x f'(y)dy$$

thanks for uniqueness :

1. let u_1, u_2 solve ✓
2. define $v = u_1 - u_2$. by linearity this means v solves as well, but for $f(x) = 0$ ✓
3. apply stability/energy estimate to show $v = 0$ ✓

Taking an absolute value of both sides yields

$$|f(x)| = \left| \int_0^x f'(y)dy \right| \leq \int_0^x |f'(y)|dy \leq \int_0^1 |f'(y)|dy$$

Squaring both sides yields:

$$|f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy \right)^2$$

I claim that

$$\left(\int_0^1 |f'(y)|dy \right)^2 \leq \int_0^1 |f'(y)|^2 dy.$$

Note that for two numbers a, b, we have

$$\left(\frac{1}{2}a + \frac{1}{2}b \right)^2 \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$$

because $x \mapsto x^2$ is convex. For a Riemann sum at $x_j = j/2^N$ of some function g, we have

$$\left(\sum_{j=1}^{2^N} \frac{1}{2^N} g(x_i) \right)^2 \leq \frac{1}{2} \left(\sum_{j=1}^{2^{N-1}} g(x_j) \right)^2 + \frac{1}{2} \left(\sum_{j=2^{N-1}+1}^{2^N} \frac{1}{2^{N-1}} g(x_j) \right)^2.$$

We can continue recursively to show

$$\left(\frac{1}{2^N} \sum_{j=1}^{2^N} g(x_i) \right)^2 \leq \frac{1}{2^N} \sum_{j=1}^{2N} g(x_i)^2$$

The limit of these sums as $N \to \infty$ is

$$\left(\int_0^1 g(x)dx \right)^2 \leq \int_0^1 g(x)^2 dx$$

Setting $g(x) = |f'(x)|$ proves the claim. Note that this is a special case of **Jensen's inequality**. We now insert the claim into our inequality

$$|f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy \right)^2 \leq \int_0^1 |f'(y)|^2 dy$$

Taking the max over all $x \in [0, 1]$ yields

$$\max_{x \in [0,1]} |f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy \right)^2 \leq \int_0^1 |f'(y)|^2 dy$$

and

$$\int_0^1 |f(x)|^2 dx \leq \max_{x\in[0,1]} |f(x)|^2 \leq \int_0^1 |f'(x)|^2 dx.$$

□

Remark 3.1 (another version of Poincare)**.** Note that we technically proved a stronger version of Poincare:

$$\max_{x\in[0,1]} |f(x)|^2 \leq \int_0^1 |f'(x)|^2 dx.$$

This can be used to prove an alternative stability result for Poisson's equation without using maximum principle.

COMPUTATIONAL PDE LECTURE 9

LUCAS BOUCK

1. Outline of today

- Prove Poincare inequality
- Begin separation of variables.

2. Energy estimates for the heat equation

Recall that last time we proved.

Proposition 2.1 (energy estimate)**.** Let u be a C^2 solution to the heat equation with homogenous Dirichlet boundary conditions.

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), & t > 0 \text{ and } x \in (0,1) \\ u(0) = u(1) = 0, & \text{(boundary condition)} \\ u(0,x) = u_0(x) & \text{(initial condition)} \end{cases} \tag{1}$$

Then, for all $T > 0$, we have

$$\int_0^1 |u(T,x)|^2 dx + \int_0^T \int_0^1 |u_x(t,x)|^2 dx dt \leq \int_0^1 |u_0(x)|^2 dx + \int_0^T \int_0^1 |f(t,x)|^2 dx dt \tag{2}$$

A corollary of the energy estimate is that solutions to the heat equation are unique.

Corollary 2.1 (uniqueness of solutions)**.** C^2 solutions to (1) are unique.

Proof. Let u_1, u_2 be C^2 solutions to (1). Then the difference $v = u_1 - u_2$ solves

$$\begin{cases} v_t(t,x) - v_{xx}(t,x) = 0, & t > 0 \text{ and } x \in (0,1) \\ v(0) = v(1) = 0, \\ v(0,x) = 0 \end{cases}$$

Applying the energy estimates to v shows $v = 0$. □

Lemma 2.1 (Poincare's inequality)**.** Let $f \in C^1[0,1]$ satisfy $f(0) = 0$, then

$$\int_0^1 |f(x)|^2 dx \leq \int_0^1 |f'(x)|^2 dx$$

with $f(0) = 0$

WTS $\int_0^1 |f(x)|^2 dx \leq \int_0^1 |f'(x)|^2 dx$

Date: September 18, 2023.

$f(x) - f(0) = \int_0^x f'(y)dy$

$\hookrightarrow f(x) = 0 + \int_0^x f'(y)dy$

Proof. We can write

$$f(x) = \cancel{f(0)} + \int_0^x f'(y)dy$$

Taking an absolute value of both sides yields

$$|f(x)| = \left|\int_0^x f'(y)dy\right| \leq \int_0^x |f'(y)|dy \leq \int_0^1 |f'(y)|dy$$

Squaring both sides yields:

$$|f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy\right)^2$$

I claim that

jensen's inequality

$$\left(\int_0^1 |f'(y)|dy\right)^2 \leq \int_0^1 |f'(y)|^2 dy.$$

Note that for two numbers a, b, we have

$$\left(\frac{1}{2}a + \frac{1}{2}b\right)^2 \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$$

because $x \mapsto x^2$ is convex. For a Riemann sum at $x_j = j/2^N$ of some function g, we have

$$\left(\sum_{j=1}^{2^N} \frac{1}{2^N} g(x_i)\right)^2 \leq \frac{1}{2}\left(\sum_{j=1}^{2^{N-1}} g(x_j)\right)^2 + \frac{1}{2}\left(\sum_{j=2^{N-1}+1}^{2^N} \frac{1}{2^{N-1}} g(x_j)\right)^2.$$

We can continue recursively to show

$$\left(\frac{1}{2^N}\sum_{j=1}^{2^N} g(x_i)\right)^2 \leq \frac{1}{2^N}\sum_{j=1}^{2N} g(x_i)^2$$

The limit of these sums as $N \to \infty$ is

$$\left(\int_0^1 g(x)dx\right)^2 \leq \int_0^1 g(x)^2 dx$$

Setting $g(x) = |f'(x)|$ proves the claim. Note that this is a special case of **Jensen's inequality**. We now insert the claim into our inequality

$$|f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy\right)^2 \leq \int_0^1 |f'(y)|^2 dy$$

Stronger version of poincare : $\int_0^1 |f(x)|^2 dx \leq \max_{x\in[0,1]} |f(x)|^2$

Taking the max over all $x \in [0,1]$ yields

$$\max_{x\in[0,1]} |f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy\right)^2 \leq \int_0^1 |f'(y)|^2 dy$$

and

$$\int_0^1 |f(x)|^2 dx \leq \max_{x\in[0,1]} |f(x)|^2 \leq \int_0^1 |f'(x)|^2 dx.$$

□

Remark 2.1 (another version of Poincare)**.** Note that we technically proved a stronger version of Poincare:

$$\max_{x\in[0,1]} |f(x)|^2 \leq \int_0^1 |f'(x)|^2 dx.$$

This can be used to prove an alternative stability result for Poisson's equation without using maximum principle.

3. Separation of variables

We now construct solutions to the heat equation. We start with

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = 0, \quad t > 0, x \in (0,1) \\ u(t,0) = u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases} \tag{3}$$

The main idea of separation of variables is to look for solutions of the form

$$u(t,x) = \sum_{k=1}^{N} c_k u_k(t,x)$$

where we can separate the variables of each u_k:

$$u_k(t,x) = T_k(t)X_k(x).$$

Importantly, if each u_k solves the differential equation in (3), i.e.

$$\partial_t u_k(t,x) - \partial_{xx}^2 u_k(t,x) = 0,$$

then u solves the differential equation because:

$$u_t(t,x) - u_{xx}(t,x) = \sum_{k=1}^{N} c_k \partial_t u_k(t,x) - \partial_{xx}^2 u_k(t,x) = 0.$$

We have used the fact that the differential equation is linear. This is also known as the **Principle of Superposition**.

We now try to solve for each u_k. We plug in $u_k(t,x) = T_k(t)X_k(x)$ into the differential equation to get

$$\partial_t u_k(t,x) - \partial_{xx}^2 u_k(t,x) = T_k'(t)X_k(x) - T_k(t)X_k''(x) = 0.$$

Rearranging the equation leads to

$$\frac{T_k'(t)}{T_k(t)} = \frac{X_k''(x)}{X_k(x)}.$$

Notice that each side of the above equation is equal for all choices of t, x. The only functions that can satisfy this are constant functions. We'll let λ_k denote the constant such that

$$\frac{T_k'(t)}{T_k(t)} = \frac{X_k''(x)}{X_k(x)} = \lambda_k,$$

and we now solve for X_k and T_k separately.

We first solve for X such that

$$X_k''(x) = \lambda_k X_k(x), \quad X_k(0) = X_k(1) = 0.$$

Note that we enforce the boundary condition on X_k in order for u_k and u to satisfy the boundary conditions. We saw in HW 1, that the possible solutions to this problem (up to a multiplying constant) are

$$X_k(x) = \sin(k\pi x), \quad \lambda_k = -k^2\pi^2.$$

We can also solve for

$$T_k'(t) = k^2\pi^2 T_k(t),$$

whose solutions are

$$T_k(t) = T_k(0)e^{-k^2\pi^2 t}.$$

Hence, our candidate solution u is

$$u(t,x) = \sum_{k=1}^{N} c_k T_k(t)X_k(x) = \sum_{k=1}^{N} c_k e^{-k^2\pi^2 t}\sin(k\pi x)$$

where c_k are constants that are to be determined, which we will do next time.

COMPUTATIONAL PDE LECTURE 9

LUCAS BOUCK

1. Outline of today

- Prove Poincare inequality
- Begin separation of variables.

2. Energy estimates for the heat equation

Recall that last time we proved.

Proposition 2.1 (energy estimate)**.** Let u be a C^2 solution to the heat equation with homogenous Dirichlet boundary conditions.

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), & t > 0 \text{ and } x \in (0,1) \\ u(0) = u(1) = 0, & \text{(boundary condition)} \\ u(0,x) = u_0(x) & \text{(initial condition)} \end{cases} \tag{1}$$

Then, for all $T > 0$, we have

$$\int_0^1 |u(T,x)|^2 dx + \int_0^T \int_0^1 |u_x(t,x)|^2 dx dt \leq \int_0^1 |u_0(x)|^2 dx + \int_0^T \int_0^1 |f(t,x)|^2 dx dt \tag{2}$$

A corollary of the energy estimate is that solutions to the heat equation are unique.

Corollary 2.1 (uniqueness of solutions)**.** C^2 solutions to (1) are unique.

Proof. Let u_1, u_2 be C^2 solutions to (1). Then the difference $v = u_1 - u_2$ solves

$$\begin{cases} v_t(t,x) - v_{xx}(t,x) = 0, & t > 0 \text{ and } x \in (0,1) \\ v(0) = v(1) = 0, \\ v(0,x) = 0 \end{cases}$$

Applying the energy estimates to v shows $v = 0$. □

Lemma 2.1 (Poincare's inequality)**.** Let $f \in C^1[0,1]$ satisfy $f(0) = 0$, then

$$\int_0^1 |f(x)|^2 dx \leq \int_0^1 |f'(x)|^2 dx$$

Date: September 18, 2023.

Proof. We can write

$$f(x) = \cancel{f(0)} + \int_0^x f'(y)dy$$

Taking an absolute value of both sides yields

$$|f(x)| = \left|\int_0^x f'(y)dy\right| \leq \int_0^x |f'(y)|dy \leq \int_0^1 |f'(y)|dy$$

Squaring both sides yields:

$$|f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy\right)^2$$

I claim that

$$\left(\int_0^1 |f'(y)|dy\right)^2 \leq \int_0^1 |f'(y)|^2 dy.$$

Note that for two numbers a, b, we have

$$\left(\frac{1}{2}a + \frac{1}{2}b\right)^2 \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$$

because $x \mapsto x^2$ is convex. For a Riemann sum at $x_j = j/2^N$ of some function g, we have

$$\left(\sum_{j=1}^{2^N} \frac{1}{2^N} g(x_i)\right)^2 \leq \frac{1}{2}\left(\sum_{j=1}^{2^{N-1}} g(x_j)\right)^2 + \frac{1}{2}\left(\sum_{j=2^{N-1}+1}^{2^N} \frac{1}{2^{N-1}} g(x_j)\right)^2.$$

We can continue recursively to show

$$\left(\frac{1}{2^N}\sum_{j=1}^{2^N} g(x_i)\right)^2 \leq \frac{1}{2^N}\sum_{j=1}^{2N} g(x_i)^2$$

The limit of these sums as $N \to \infty$ is

$$\left(\int_0^1 g(x)dx\right)^2 \leq \int_0^1 g(x)^2 dx$$

Setting $g(x) = |f'(x)|$ proves the claim. Note that this is a special case of **Jensen's inequality**. We now insert the claim into our inequality

$$|f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy\right)^2 \leq \int_0^1 |f'(y)|^2 dy$$

Taking the max over all $x \in [0,1]$ yields

$$\max_{x\in[0,1]} |f(x)|^2 \leq \left(\int_0^1 |f'(y)|dy\right)^2 \leq \int_0^1 |f'(y)|^2 dy$$

and

$$\int_0^1 |f(x)|^2 dx \leq \max_{x\in[0,1]} |f(x)|^2 \leq \int_0^1 |f'(x)|^2 dx.$$

□

Remark 2.1 (another version of Poincare)**.** Note that we technically proved a stronger version of Poincare:

$$\max_{x\in[0,1]} |f(x)|^2 \leq \int_0^1 |f'(x)|^2 dx.$$

This can be used to prove an alternative stability result for Poisson's equation without using maximum principle.

3. Separation of variables

We now construct solutions to the heat equation. We start with

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = 0, \quad t > 0, x \in (0,1) \\ u(t,0) = u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases} \tag{3}$$

The main idea of separation of variables is to look for solutions of the form

$$u(t,x) = \sum_{k=1}^{N} c_k u_k(t,x)$$

where we can separate the variables of each u_k:

$$u_k(t,x) = T_k(t)X_k(x).$$

Importantly, if each u_k solves the differential equation in (3), i.e.

$$\partial_t u_k(t,x) - \partial_{xx}^2 u_k(t,x) = 0,$$

then u solves the differential equation because:

$$u_t(t,x) - u_{xx}(t,x) = \sum_{k=1}^{N} c_k \partial_t u_k(t,x) - \partial_{xx}^2 u_k(t,x) = 0.$$

We have used the fact that the differential equation is linear. This is also known as the **Principle of Superposition**.

separation of variables: look for solns in the following form

$u(t,x) = \sum_{k=1}^{N} c_k u_k(t,x)$

where

$u_k(t,x) = T_k(t) X_k(x)$

if each u_k solves, then by linearity the entire u also solves

We now try to solve for each u_k. We plug in $u_k(t,x) = T_k(t)X_k(x)$ into the differential equation to get

$$\partial_t u_k(t,x) - \partial_{xx}^2 u_k(t,x) = T_k'(t)X_k(x) - T_k(t)X_k''(x) = 0.$$

Rearranging the equation leads to

$$\frac{T_k'(t)}{T_k(t)} = \frac{X_k''(x)}{X_k(x)}.$$

Notice that each side of the above equation is equal for all choices of t, x. The only functions that can satisfy this are constant functions. We'll let λ_k denote the constant such that

$$\frac{T_k'(t)}{T_k(t)} = \frac{X_k''(x)}{X_k(x)} = \lambda_k,$$

and we now solve for X_k and T_k separately.

We first solve for X such that

$$X_k''(x) = \lambda_k X_k(x), \quad X_k(0) = X_k(1) = 0.$$

Note that we enforce the boundary condition on X_k in order for u_k and u to satisfy the boundary conditions. We saw in HW 1, that the possible solutions to this problem (up to a multiplying constant) are

$$X_k(x) = \sin(k\pi x), \quad \lambda_k = -k^2\pi^2.$$

We can also solve for

$$T_k'(t) = k^2\pi^2 T_k(t),$$

whose solutions are

$$T_k(t) = T_k(0)e^{-k^2\pi^2 t}.$$

Hence, our candidate solution u is

$$u(t,x) = \sum_{k=1}^{N} c_k T_k(t)X_k(x) = \sum_{k=1}^{N} c_k e^{-k^2\pi^2 t}\sin(k\pi x)$$

where c_k are constants that are to be determined, which we will do next time.

COMPUTATIONAL PDE LECTURE 10

LUCAS BOUCK

1. Outline of today

- Continue separation of variables

2. Separation of Variables

Last time we have considered:

$$(1) \qquad \begin{cases} u_t(t,x) - u_{xx}(t,x) = 0, & t > 0, x \in (0,1) \\ u(t,0) = u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases}$$

and have shown that

$$u(t,x) = \sum_{k=1}^{N} c_k T_k(t) X_k(x)$$

solves the boundary conditions and differential equation in (1) where

$$T_k(t) = e^{-(k\pi)^2 t}, \quad X_k(t) = \sin(k\pi x).$$

Additionally, if

$$u_0(x) = \sum_{k=1}^{N} c_k \sin(k\pi x),$$

then u also satisfies the initial condition.

We now address what if u_0 is not the finite sum of sines.

2.1. More general initial conditions: Fourier series.

Suppose

$$u_0(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x),$$

then

$$u(t,x) = \sum_{k=1}^{\infty} c_k T_k(t) X_k(x)$$

Date: September 20, 2023.

would formally solve (1). I say formal solution because we have not actually proved that the infinite series makes sense. We now show a sufficient condition for the infinite series to make sense.

Lemma 2.1 (Weierstrass M-test)**.** Let $f_k : [0,1] \to \mathbb{R}$ be a sequence of continuous functions with $M_k = \max_{x\in[0,1]} |f_k(x)|$. If

$$\sum_{k=1}^{\infty} M_k < \infty,$$

then for any x, the series

$$\sum_{k=1}^{\infty} f_k(x)$$

converges absolutely. Also, define $f(x) = \sum_{k=1}^{\infty} f_k(x)$. We also have that the series converges uniformly, i.e.

$$\lim_{N\to\infty} \max_{x\in[0,1]} \left| f(x) - \sum_{k=1}^{N} f_k(x) \right| = 0,$$

and f is continuous.

Proof. We show that the sequence $F_N = \sum_{k=1}^{N} f_k$ is uniformly Cauchy. Let $m \geq N$, and we compute

$$\left|F_N(x) - F_m(x)\right| \leq \sum_{k=N+1}^{m} |f_k(x)| \leq \sum_{k=N+1}^{m} M_k \leq \sum_{k=N+1}^{\infty} M_k$$

Taking a max over all x leads to

$$\max_{x\in[0,1]} \left|F_N(x) - F_m(x)\right| \leq \sum_{k=N+1}^{\infty} M_k$$

and taking a limit as $N \to \infty$ shows

$$\lim_{N\to\infty} \sup_{m\geq N} \max_{x\in[0,1]} \left|F_N(x) - F_m(x)\right| \leq 0$$

Hence, F_N is uniformly Cauchy, and converges uniformly to some continuous f. □

The relevant result for us is the following

Proposition 2.1. Let $\{c_k\}_{k\in\mathbb{N}}$ be a sequence and define:

$$u_0^N(x) = \sum_{k=1}^{N} c_k \sin(k\pi x)$$

If

$$\sum_{k=1}^{\infty} |c_k| < \infty,$$

then u_0^N converges uniformly on $[0,1]$ and the limit is the continuous function

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

ie. the infinite series makes sense

2.1.1. *Computing the coefficients.* Suppose u_0 is some continuous function. How do we find c_k such that

$$u_0(x) = \sum_{k=1}^{\infty} c_k \sin(k\pi x)?$$

this has to do w. find the coeffs ?

The coefficients in this case are known as **Fourier coefficients**.

We first begin by looking at linear algebra. Given a vector $\mathbf{w} \in \mathbb{R}^n$ what is the best approximation $\mathbf{v}^* \in \mathbb{V}$, where $\mathbb{V} \subset \mathbb{R}^n$ is a subset of $\mathbb{R}^n$. We first consider the Euclidean norm

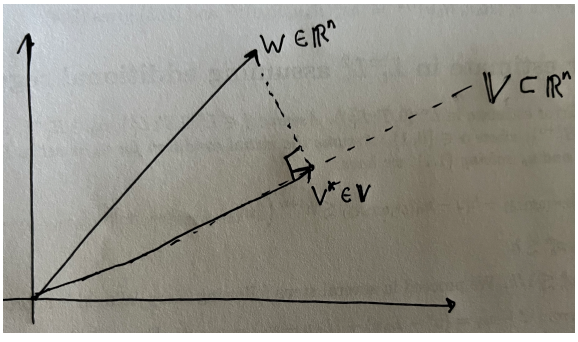
$$\|\mathbf{v}\|_2 = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

Then

$$\|\mathbf{v}^* - \mathbf{w}\|_2 = \inf_{\mathbf{u} \in \mathbb{V}} \|\mathbf{u} - \mathbf{w}\|_2$$

if and only if

$$(\mathbf{v}^* - \mathbf{w}) \cdot \mathbf{u} = 0 \text{ for all } \mathbf{u} \in \mathbb{V}.$$



To see this one direction of the if and only if, we can write

$$\begin{aligned}
\|\mathbf{v}^* - \mathbf{w}\|_2^2 &= \mathbf{v}^* \cdot (\mathbf{v}^* - \mathbf{w}) - \mathbf{w} \cdot (\mathbf{v}^* - \mathbf{w}) \\
&= -\mathbf{w} \cdot (\mathbf{v}^* - \mathbf{w}) = (\mathbf{u} - \mathbf{w}) \cdot (\mathbf{v}^* - \mathbf{w}) \\
&\leq \|(\mathbf{u} - \mathbf{w})\|_2 \|(\mathbf{v}^* - \mathbf{w})\|_2 \qquad \text{(Cauchy-Schwarz inequality)}
\end{aligned}$$

Dividing both sides by $\|(\mathbf{v}^* - \mathbf{w})\|_2$ shows that

$$\|\mathbf{v}^* - \mathbf{w}\|_2 \leq \inf_{\mathbf{u} \in \mathbb{V}} \|\mathbf{u} - \mathbf{w}\|_2.$$

The vector $\mathbf{v}^*$ is known as the **orthogonal projection** of $\mathbf{w}$ onto $\mathbb{V}$. Suppose $\{\mathbf{v}^k\}_{k=1}^m$ is an orthogonal basis for $\mathbb{V}$. That is suppose $\{\mathbf{v}^k\}_{k=1}^m$ is a basis for $\mathbb{V}$ and $\mathbf{v}^k \cdot \mathbf{v}^j = 0$ for all $j \neq k$. Then, we write

$$\mathbf{v}^* = \sum_{k=1}^m a_k \mathbf{v}^k$$

and subtract $\mathbf{w}$ and take a dot product with $\mathbf{v}^j$ to get

$$\begin{aligned} 0 = (\mathbf{v}^* - \mathbf{w}) \cdot \mathbf{v}^j &= \left(\sum_{k=1}^m a_k \mathbf{v}^k \right) \cdot \mathbf{v}^j - \mathbf{w} \cdot \mathbf{v}^j \\ &= a_k \|\mathbf{v}^j\|_2^2 - \mathbf{w} \cdot \mathbf{v}^j, \end{aligned}$$

and

$$a_k = \frac{\mathbf{w} \cdot \mathbf{v}^j}{\|\mathbf{v}^j\|_2^2}.$$

For us, we need to somehow mimic the dot product, but for functions. The relevant inner product is the L^2 inner product:

$$\langle u, v \rangle = \int_0^1 u(x)v(x)dx.$$

and L^2 norm:

$$\|u\|_{L^2[0,1]} = \sqrt{\langle u, u \rangle}.$$

Luckily, our sine basis is orthogonal in the L^2 inner product.

Lemma 2.2. Let $X_k(x) = \sin(\pi k x)$. Then

$$\langle X_k, X_j \rangle = \begin{cases} 0, & j \neq k \\ 1/2, & j = k \end{cases}.$$

We then have the following approximation result.

Proposition 2.2. Let $X_k(x) = \sin(\pi k x)$. Let u_0 be continuous. Then, we define the Fourier coefficient as

$$c_k = \frac{\langle X_k, X_j \rangle}{\langle X_k, X_k \rangle}$$

and the sum

$$u_0^N(x) = \sum_{k=1}^N c_k X_k$$

is the best L^2 approximation of u_0 in $\mathbb{V} = \text{span}\{X_k\}_{k=1}^N$. That is,

$$\int_0^1 |u_0(x) - u_0^N(x)|^2 dx = \inf_{v \in \mathbb{V}} \int_0^1 |u_0(x) - v(x)|^2 dx.$$

Moreover if the coefficients c_k satisfy

$$\sum_{k=1}^{\infty} |c_k| < \infty,$$

then $u_0^N \to u_0$ uniformly.

Remark 2.1 (estimates on Fourier coefficients)**.** The proof of this can be done using integration by parts.

If $u_0 \in C^2(0,1)$ with $u_0(0) = u_0(1) = u_0'(0) = u_0'(1) = 0$, then

$$|c_k| \leq \frac{\max_{x \in [0,1]} |u_0''(x)|}{k^2 \pi^2}$$

and

$$\sum_{k=1}^{\infty} |c_k| < \infty.$$

Ultimately, to compute a solution to the heat equation (1), we follow the following procedure.

- Compute the Fourier coefficients

$$c_k = 2 \int_0^1 u_0(x) \sin(\pi k x) dx$$

- Verify that

$$\sum_{k=1}^{\infty} |c_k| < \infty$$

- Write down

$$u(t,x) = \sum_{k=1}^{\infty} c_k e^{-(k\pi)^2 t} \sin(k\pi x)$$

and u solves (1).

if u solves

$$\begin{cases} u_t - u_{xx} = 0 \\ u(t,0) = u(t,1) = 0 \\ u(x,0) = u_0(x) \end{cases}$$

easiest case. we can compute this directly

to compute soln to heat eqn :

1. compute $c_k = 2\int_0^1 u_0(x)\sin(k\pi x)dx$

2. verify that $\sum_{k=1}^{\infty} |c_k| < \infty$

3. write $u(t,x) = \sum_{k=1}^{\infty} c_k e^{-(k\pi)^2 t} \sin(k\pi x)$

COMPUTATIONAL PDE LECTURE 11

LUCAS BOUCK

1. Outline of today

- Explain separation of variables for different boundary conditions and right hand side

2. Separation of Variables

Last time we have considered:

$$\text{(1)} \qquad \begin{cases} u_t(t,x) - u_{xx}(t,x) = 0, \quad t > 0, x \in (0,1) \\ u(t,0) = u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases}$$

and have shown that

$$u(t,x) = \sum_{k=1}^{\infty} c_k T_k(t) X_k(x)$$

solves the boundary conditions and differential equation in (1) where

$$T_k(t) = e^{-(k\pi)^2 t}, \quad X_k(t) = \sin(k\pi x).$$

and

$$c_k = \frac{\langle u_0, X_k \rangle}{\langle X_k, X_k \rangle} = \frac{\int_0^1 u_0(x)\sin(k\pi x)dx}{\int_0^1 \sin^2(k\pi x)dx} = 2\int_0^1 u_0(x)\sin(k\pi x)dx.$$

We now address what happens with more general situations.

2.1. Right hand side f:

Suppose we need to solve.

$$\text{(2)} \qquad \begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), \quad t > 0, x \in (0,1) \\ u(t,0) = u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases}$$

what if $f \neq 0$?

We can reduce this into solving the 2 problems.

- w solves (2) with $f = 0$ and $u_0 \neq 0$.
- v solves (2) with $u_0 = 0$ and $f \neq 0$.

} split into two problems

Date: September 22, 2023.

$u = v + w$

v solves $f \neq 0$ w/ $u_0 = 0$

solves w/ $f = 0$ and $u_0 \neq 0$

we know w from above

The sum $u = v + w$ then solves (2). We have previously discussed how to solve for w. We now discuss how to solve for v.

- We write
$$v(t,x) = \sum_{k=1}^{\infty} T_k(t) X_k(x),$$
where $X_k(x) = \sin(k\pi x)$ from before.
- We also write the Fourier series for f:
$$f(t,x) \sum_{k=1}^{\infty} a_k(t) X_k(x).$$
- The heat equation now looks like
$$\sum_{k=1}^{\infty} (T_k'(t) + \pi^2 k^2 T_k(t)) X_k(x) = \sum_{k=1}^{\infty} a_k(t) X_k(x)$$
- We then solve each ODE initial value problem:
$$T_k'(t) + \pi^2 k^2 T_k(t) = a_k(t), \quad T_k(t) = 0$$
separately.

Using the example from recitation, if
$$f(t,x) = \sin(\pi x),$$
then $a_1(t) = 1$, and $a_k(t) = 0$ for all $k > 1$. Then,
$$T_1(t) = \frac{1}{\pi^2}\left(1 - e^{-\pi^2 t}\right),$$
and
$$v(t,x) = \frac{1}{\pi^2}\left(1 - e^{-\pi^2 t}\right)\sin(\pi x)$$

2.2. **Different boundary conditions.** There are two separate cases we will consider with different boundary conditions.

2.2.1. *Nonhomogenous boundary conditions.* Consider the situation where

$$\text{(3)} \qquad \begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), & t > 0, x \in (0,1) \\ u(t,0) = u_\ell(t), u(t,1) = u_r(t) \\ u(0,x) = u_0(x) \end{cases}$$

with $u_0(0) = u_\ell(0)$ and $u_0(1) = u_r(0)$. To solve for u, we write
$$g(t,x) = u_\ell(t)(1-x) + u_r(t)x$$

and solve for $v = u - g$, where v solves

$$\begin{cases} v_t(t,x) - v_{xx}(t,x) = f(t,x) - g_t(t,x) + g_{xx}(t,x), & t > 0, x \in (0,1) \\ v(t,0) = 0, u(t,1) = 0 \\ v(0,x) = u_0(x) - g(0,x) \end{cases}$$

and apply the procedure from the previous section to solve for v. Then $u = v + g$ is the solution to (3).

2.2.2. *Neumann boundary conditions.* This section address what happens when we have homogenous Neumann boundary conditions:

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = 0, & t > 0, x \in (0,1) \\ u_x(t,0) = 0, u_x(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases} \tag{4}$$

We again write the solution as

$$u(t,x) = \sum_{k=1}^{\infty} T_k(t)\tilde{X}_k(x)$$

where $\tilde{X}_k$ is a different function than $X_k(x) = \sin(k\pi x)$. This is because X_k does not satisfy the boundary conditions $X_k'(0) = X_k'(1) = 0$. Plugging u into the heat equation yields

$$\sum_{k=1}^{\infty} \left(T_k'(t)\tilde{X}_k(x) - T_k(t)\tilde{X}_k''(x) \right) = 0.$$

Solving for each individual k brings us back to

$$\frac{T_k'(t)}{T_k(t)} = \frac{X_k''(x)}{X_k(x)} = \tilde{\lambda}_k$$

where $\tilde{\lambda}_k$ is a constant. Therefore, we need to solve the eigenvalue problem

$$\begin{cases} \tilde{X}''(x) - \tilde{\lambda}\tilde{X}(x) = 0 \\ \tilde{X}'(0) = \tilde{X}'(1) = 0 \end{cases} \tag{5}$$

We now go over how to solve these eigenvalue problems.

Proposition 2.1. The values $\tilde{\lambda}_k = -(k\pi)^2$ and $\tilde{X}_k(x) = \cos(k\pi x)$ for $k = 0, \dots$ are the only nonzero solutions to the Neumann eigenvalue problem (5).

Proof. We break the proof into 2 cases.

Case 1: Suppose $\tilde{X}, \tilde{\lambda}$ solve (5) with $\tilde{\lambda} > 0$. Then we have that $\tilde{\lambda} = a^2$ for some $a > 0$. Then the solution to $\tilde{X}'' = a^2\tilde{X}$ is

$$\tilde{X}(x) = c_1 e^{ax} + c_2 e^{-ax}.$$

Notice that $\tilde{X}'(0) = c_1 a - c_2 a$, so in order to satisfy $\tilde{X}'(0) = 0$, we require $c_1 = c_2$. If this is the case, then $\tilde{X}'(0) = c_1 a e^a - c_1 a e^{-a} = c_1 a(e^a - e^{-a})$. The only way for $c_1 a(e^a - e^{-a}) = 0$ for $a > 0$ is if $c_1 = 0$. Thus, $\tilde{X}(x) = 0$. Hence there cannot be a nonzero eigenfunction with $\tilde{\lambda} > 0$.

Case 2: Suppose $\tilde{X}, \tilde{\lambda}$ solve (5) with $\tilde{\lambda} \leq 0$. Then we have that $\tilde{\lambda} = -a^2$ for some $a \geq 0$. The solution to $\tilde{X}'' = -a^2 \tilde{X}$ is

$$\tilde{X}(x) = c_1 \cos(ax) + c_2 \sin(ax).$$

To satisfy the boundary conditions, we require

$$0 = \tilde{X}'(0) = \cancel{-ac_1 \sin(a0)} + ac_2 \cos(a0) = ac_2$$
$$0 = \tilde{X}'(0) = -ac_1 \sin(a) + ac_2 \cos(a)$$

To solve the first equation, we require $a = 0$ or $c_2 = 0$.

If $a = 0$, then

$$\tilde{X}(x) = c_1$$

is a constant. Since we do not care about the constant scaling of an eigenvector, we take $c_1 = 1$.

If $a > 0$, then $c_2 =$, and we require $-ac_1 \sin(a) = 0$. In order for $\tilde{X} \neq 0$, we need $\sin(a) = 0$, and the only solutions for $a > 0$ are $a = \pi k$ for $k = 1, \ldots$.

In conclusion, the only nonzero $\tilde{X}$ to solve (5) are $\tilde{\lambda}_k = -(k\pi)^2$ and $\tilde{X}_k(x) = \cos(k\pi x)$ for $k = 0, \ldots$, which concludes the proof. □

This same procedure of breaking the eigenvalue into cases $\lambda > 0$ and $\lambda \leq 0$ can also show the following

Proposition 2.2. The values $\lambda_k = -(k\pi)^2$ and $X_k(x) = \sin(k\pi x)$ for $k = 1, \ldots$ are the only nonzero solutions to the Dirichlet eigenvalue problem:

$$\begin{cases} X''(x) - \lambda X(x) = 0 \\ X(0) = X(1) = 0 \end{cases}. \tag{6}$$

Proof. This will be a homework problem. □

COMPUTATIONAL PDE LECTURE 12

LUCAS BOUCK

1. Outline of today

- Finish separation of variables
- Start finite differences

2. Neumann boundary conditions

This section address what happens when we have homogenous Neumann boundary conditions:

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = 0, \quad t > 0, x \in (0,1) \\ u_x(t,0) = 0, u_x(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

We again write the solution as

$$u(t,x) = \sum_{k=1}^{\infty} T_k(t)\tilde{X}_k(x)$$

where $\tilde{X}_k$ is a different function than $X_k(x) = \sin(k\pi x)$. This is because X_k does not satisfy the boundary conditions $X_k'(0) = X_k'(1) = 0$. Plugging u into the heat equation yields

$$\sum_{k=1}^{\infty} \left(T_k'(t)\tilde{X}_k(x) - T_k(t)\tilde{X}_k''(x) \right) = 0.$$

Solving for each individual k brings us back to

$$\frac{T_k'(t)}{T_k(t)} = \frac{X_k''(x)}{X_k(x)} = \tilde{\lambda}_k$$

where $\tilde{\lambda}_k$ is a constant. Therefore, we need to solve the eigenvalue problem

$$\begin{cases} \tilde{X}''(x) - \tilde{\lambda}\tilde{X}(x) = 0 \\ \tilde{X}'(0) = \tilde{X}'(1) = 0 \end{cases}. \tag{2}$$

We now go over how to solve these eigenvalue problems.

Date: September 25, 2023.

Proposition 2.1. The values $\tilde{\lambda}_k = -(k\pi)^2$ and $\tilde{X}_k(x) = \cos(k\pi x)$ for $k = 0, \dots$ are the only nonzero solutions to the Neumann eigenvalue problem (2).

Proof. We break the proof into 2 cases.

Case 1: Suppose $\tilde{X}, \tilde{\lambda}$ solve (2) with $\tilde{\lambda} > 0$. Then we have that $\tilde{\lambda} = a^2$ for some $a > 0$. Then the solution to $\tilde{X}'' = a^2\tilde{X}$ is

$$\tilde{X}(x) = c_1 e^{ax} + c_2 e^{-ax}.$$

Notice that $\tilde{X}'(0) = c_1 a - c_2 a$, so in order to satisfy $\tilde{X}'(0) = 0$, we require $c_1 = c_2$. If this is the case, then $\tilde{X}'(0) = c_1 a e^a - c_1 a e^{-a} = c_1 a(e^a - e^{-a})$. The only way for $c_1 a(e^a - e^{-a}) = 0$ for $a > 0$ is if $c_1 = 0$. Thus, $\tilde{X}(x) = 0$. Hence there cannot be a nonzero eigenfunction with $\tilde{\lambda} > 0$.

Case 2: Suppose $\tilde{X}, \tilde{\lambda}$ solve (2) with $\tilde{\lambda} \leq 0$. Then we have that $\tilde{\lambda} = -a^2$ for some $a \geq 0$. The solution to $\tilde{X}'' = -a^2\tilde{X}$ is

$$\tilde{X}(x) = c_1 \cos(ax) + c_2 \sin(ax).$$

To satisfy the boundary conditions, we require

$$0 = \tilde{X}'(0) = \cancel{-ac_1 \sin(a0)} + ac_2 \cos(a0) = ac_2$$
$$0 = \tilde{X}'(0) = -ac_1 \sin(a) + ac_2 \cos(a)$$

To solve the first equation, we require $a = 0$ or $c_2 = 0$.

If $a = 0$, then

$$\tilde{X}(x) = c_1$$

is a constant. Since we do not care about the constant scaling of an eigenvector, we take $c_1 = 1$.

If $a > 0$, then $c_2 =$, and we require $-ac_1 \sin(a) = 0$. In order for $\tilde{X} \neq 0$, we need $\sin(a) = 0$, and the only solutions for $a > 0$ are $a = \pi k$ for $k = 1, \dots$.

In conclusion, the only nonzero $\tilde{X}$ to solve (2) are $\tilde{\lambda}_k = -(k\pi)^2$ and $\tilde{X}_k(x) = \cos(k\pi x)$ for $k = 0, \dots$, which concludes the proof. □

This same procedure of breaking the eigenvalue into cases $\lambda > 0$ and $\lambda \leq 0$ can also show the following

Proposition 2.2. The values $\lambda_k = -(k\pi)^2$ and $X_k(x) = \sin(k\pi x)$ for $k = 1, \dots$ are the only nonzero solutions to the Dirichlet eigenvalue problem:

$$\begin{cases} X''(x) - \lambda X(x) = 0 \\ X(0) = X(1) = 0 \end{cases}. \tag{3}$$

Proof. This will be a homework problem. □

Remark 2.1 (general procedure)**.** The general procedure for separation of variables for the following problem

$$\begin{cases} u_t(t,x) - au_{xx}(t,x) + bu_x(t,x) + cu(t,x) = 0, \quad t > 0, x \in (0,1) \\ \alpha u_x(t,0) + \beta u(t,0) = 0, \alpha u_x(t,1) + \beta u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases}$$

- Write

$$u(t,x) = \sum_{k=0}^{\infty} T_k(t) X_k(x).$$

- Solve Eigenvalue problem

$$\begin{cases} -aX_k''(x) + bX_k'(x) + cX_k(x) - \lambda_k X_k(x) = 0, \quad t > 0, x \in (0,1) \\ \alpha X_k'(0) + \beta X_k(0) = 0, \alpha X_k'(1) + \beta X_k(1) = 0 \end{cases}$$

- Compute Fourier coefficients of u_0:

$$u_0(x) = \sum_{k=0}^{\infty} c_k X_k(x), \quad c_k = \frac{\langle u_0, X_k \rangle}{\langle X_k, X_k \rangle} = \frac{\int_0^1 u_0(x) X_k(x) dx}{\int_0^1 |X_k(x)|^2 dx}$$

- Solve initial value problems for T_k:

$$T_k'(t) = \lambda_k T_k(t), \quad T_k(0) = c_k$$

Remark 2.2 (spectral methods)**.** There is a class of numerical methods, called spectral methods, that build off of separation of variables. The idea is as follows

- Write the approximate solution as

$$U^N(t,x) = \sum_{k=0}^{N} T_k(t) X_k(x)$$

  where X_k is the basis from our eigenvalue problem
- Approximate the initial condition

$$u_0^N(x) = \sum_{k=0}^{N} \tilde{c}_k X_k(x)$$

- Use a time stepping method to solve the system of ODEs for T_k.

If the underlying solution is smooth, i.e. $u \in C^\infty$ has infinitely many continuous derivatives, we can expect *exponential convergence* of the method. That is, the truncation error is

$$\boldsymbol{\tau}^N = \mathcal{O}(e^{-N})$$

Additionally, the coefficients $\tilde{c}_k$ can be computed with a Fast Fourier Transform in $\mathcal{O}(N \log N)$ time (compared with solving a linear system, which is $\mathcal{O}(N^3)$ time).

Spectral methods are very efficient and accurate if the underlying solution is smooth.

COMPUTATIONAL PDE LECTURE 13

LUCAS BOUCK

1. Outline of today

- Start finite differences for the heat equation.

2. Finite Difference Methods for the Heat Equation

$$(1) \qquad \begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), \quad t \in (0,1), x \in (0,1) \\ u(t,0) = 0, u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases}$$

We now begin the discussion of finite difference methods for the heat equation. We begin with the setup:

- $x_j = jh$ spatial grid points $h = 1/N$
- $t_j = j\tau$ time grid points $\tau = T/M$
- Grid $G_{\tau,h} = \{t_j\}_{j=0}^M \times \{x_i\}_{i=0}^N$
- Grid function $U^{h,\tau} : G_{\tau,h} \to \mathbb{R}$

To highlight the role of time stepping, we denote $\mathbf{U}^j \in \mathbb{R}^{N-1}$ as the vector of $U^{h,\tau}$ evaluated at interior grid points, i.e.

$$\mathbf{U}_i^j = U^{h,\tau}(t_j, x_i) \text{ for } i = 1, \dots, N-1$$

and the negative second finite difference for $U^{h,\tau}(t_j, x_i)$ will be denoted by

$$\mathbf{A}^h = \begin{pmatrix} \frac{2}{h^2} & \frac{-1}{h^2} & & \\ -\frac{1}{h^2} & \frac{2}{h^2} & \frac{-1}{h^2} & \\ & \ddots & \ddots & \ddots \end{pmatrix},$$

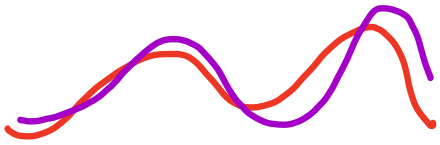
so that

$$(\mathbf{A}^h \mathbf{U}^j)_i = -D_h^2 U^{h,\tau}(t_j, x_i).$$

We finally denote the a finite difference in time as

$$D_\tau U^{h,\tau}(t_j, x_i) = \frac{U^{h,\tau}(t_j, x_i) - U^{h,\tau}(t_{j-1}, x_i)}{\tau}$$

Date: September 27, 2023.



or

$$D_\tau \mathbf{U}^j = \frac{\mathbf{U}^j - \mathbf{U}^{j-1}}{\tau}$$

2.1. **Time stepping schemes.** We now list with various time stepping schemes for the heat equation, which you implemented in recitation:

- **Forward Euler** (approximates differential equation at t_{j-1})

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^{j-1} = \mathbf{f}^{j-1}$$

- **Backward Euler** (approximates differential equation at t_j)

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^{j} = \mathbf{f}^{j}$$

- **Crank Nicolson** (approximates differential equation at $t_{j-\frac{1}{2}} = \frac{1}{2}(t_j + t_{j-1})$)

$$D_\tau \mathbf{U}^j + \frac{1}{2}\mathbf{A}^h (\mathbf{U}^j + \mathbf{U}^{j-1}) = \mathbf{f}^{j-1/2}$$

2.2. **Consistency: truncation Errors.** Recall that there are two ingredients to demonstrating convergence for finite difference schemes, which were:

- Consistency,
- Stability.

Typically it is easier to show consistency using Taylor expansion.

Proposition 2.1 (consistency of schemes)**.** Let u solve (1). Then $\mathbf{u}_i^j = u(t_j, x_i)$ satisfies

- **Forward Euler**

$$D_\tau \mathbf{u}^j + \mathbf{A}^h \mathbf{u}^{j-1} = \mathbf{f}^{j-1} + \boldsymbol{\tau}^j_{\tau,h,fe}$$

- **Backward Euler**

$$D_\tau \mathbf{u}^j + \mathbf{A}^h \mathbf{u}^{j} = \mathbf{f}^{j} + \boldsymbol{\tau}^j_{\tau,h,be}$$

- **Crank Nicolson**

$$D_\tau \mathbf{u}^j + \frac{1}{2}\mathbf{A}^h (\mathbf{u}^j + \mathbf{u}^{j-1}) = \mathbf{f}^{j+1/2} + \boldsymbol{\tau}^j_{\tau,h,cn}$$

where there is a constant $C > 0$ such that the truncation errors satisfy:

$$\|\boldsymbol{\tau}^j_{\tau,h,fe}\|_\infty \le C\left(\tau |u_{ttt}|_{C^0} + h^2 |u_{xxxx}|_{C^0}\right)$$
$$\|\boldsymbol{\tau}^j_{\tau,h,be}\|_\infty \le C\left(\tau |u_{ttt}|_{C^0} + h^2 |u_{xxxx}|_{C^0}\right)$$
$$\|\boldsymbol{\tau}^j_{\tau,h,cn}\|_\infty \le C\left(\tau^2 |u_{ttt}|_{C^0} + \tau^2 |u_{xxtt}|_{C^0} + h^2 |u_{xxxx}|_{C^0}\right)$$

and

$$|u_{ttt}|_{C^0} = \max_{t,x\in[0,1]} |u_{tt}(t,x)|, \quad |u_{ttt}|_{C^0} = \max_{t,x\in[0,1]} |u_{ttt}(t,x)|, \quad |u_{xxxx}|_{C^0} = \max_{t,x\in[0,1]} |u_{xxxx}(t,x)|$$

Proof. Recall that we have already shown:

$$\left| u_{xx}(t_j, x_i) - \frac{u(t_j, x_{i+1}) - 2u(t_j, x_i) + u(t_j, x_{i-1})}{h^2} \right| \leq Ch^2 |u_{xxxx}|_{C^0}.$$

We have seen in HW2 for a backward difference formula

$$\left| u_t(t_{j-1}, x_i) - \frac{u(t_j, x_i) - u(t_{j-1}, x_i)}{\tau} \right| \leq C\tau |u_{tt}|_{C^0}.$$

Note that we picked to evaluate the time derivative of u at t_{j-1}. For backward Euler, we'd want to look at $u_t(t_j, x_i)$. Thus, from HW2, we have the following truncation errors for each time derivative approximation

$$\textbf{Forward Euler}: \left| u_t(t_{j-1}, x_i) - \frac{u(t_j, x_i) - u(t_{j-1}, x_i)}{\tau} \right| \leq C\tau |u_{tt}|_{C^0}$$

$$\textbf{Backward Euler}: \left| u_t(t_j, x_i) - \frac{u(t_j, x_i) - u(t_{j-1}, x_i)}{\tau} \right| \leq C\tau |u_{tt}|_{C^0}$$

$$\textbf{Crank Nicolson}: \left| u_t((t_j + t_{j-1})/2, x_i) - \frac{u(t_j, x_i) - u(t_{j-1}, x_i)}{\tau} \right| \leq C\tau^2 |u_{ttt}|_{C^0}$$

The proof is complete for Forward Euler and Backward Euler.

Notice that Crank Nicolson is a centered difference approximation of u_t at the midpoint between t_j and t_{j-1}. So far we have shown for CN:

$$\begin{aligned} D_\tau u(t_j, x_i) - \frac{1}{2} D_h^2 \left(u(t_j, x_i) + u(t_{j-1}, x_i) \right) &= u_t(t_{j-1/2}, x_i) - \frac{1}{2} \left(u_{xx}(t_j, x_i) + u_{xx}(t_{j-1}, x_i) \right) + (\tilde{\boldsymbol{\tau}}^j_{h,\tau,cn})_i \\ &= u_t(t_{j-1/2}, x_i) - u_{xx}(t_{j-1/2}, x_i) + a_{i,j} + (\tilde{\boldsymbol{\tau}}^j_{h,\tau,cn})_i \\ &= f(t_{j-1/2}, x_i) + a_{i,j} + (\tilde{\boldsymbol{\tau}}^j_{h,\tau,cn})_i \end{aligned}$$

where

$$\|\tilde{\boldsymbol{\tau}}^j_{h,\tau,cn}\|_\infty \leq C\tau^2 |u_{ttt}|_{C^0} + Ch^2 |u_{xxxx}|_{C^0}$$

and

$$a_{i,j} = u_{xx}(t_{j-1/2}, x_i) - \frac{1}{2} \left(u_{xx}(t_j, x_i) + u_{xx}(t_{j-1}, x_i) \right).$$

To complete the proof of the truncation error for CN, one can show using Taylor expansion that

$$|a_{i,j}| \leq \tau^2 |u_{xxtt}|_{C^0},$$

which will be left as a HW problem. □

COMPUTATIONAL PDE LECTURE 14

LUCAS BOUCK

1. Outline of today

- Start finite differences for the heat equation.

2. Finite Difference Methods for the Heat Equation

$$(1) \qquad \begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), & t \in (0,1), x \in (0,1) \\ u(t,0) = 0, u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases}$$

2.1. **Time stepping schemes.** We now list with various time stepping schemes for the heat equation, which you implemented in recitation:

- **Forward Euler** (approximates differential equation at t_{j-1})

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^{j-1} = \mathbf{f}^{j-1}$$

- **Backward Euler** (approximates differential equation at t_j)

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^j = \mathbf{f}^j$$

- **Crank Nicolson** (approximates differential equation at $t_{j-\frac{1}{2}} = \frac{1}{2}(t_j + t_{j-1})$)

$$D_\tau \mathbf{U}^j + \frac{1}{2}\mathbf{A}^h(\mathbf{U}^j + \mathbf{U}^{j-1}) = \mathbf{f}^{j-1/2}$$

2.2. **Stability: discrete energy estimates.** The other key ingredient is stability. We'll mimic the energy estimates from the continuous heat equation. We first start with Backward Euler:

Proposition 2.1 (unconditional stability of Backward Euler)**.** Let $\mathbf{U}^j$ be the sequence of solutions to the Backward Euler scheme:

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^j = \mathbf{f}^j,$$

to show stability : mimic energy estimates

Date: September 29, 2023.

then the vector $\mathbf{U}^j$ satisfies the following discrete energy estimate:

$$\|\mathbf{U}^J\|_{2,h}^2 + \sum_{j=1}^{J} \left[\tau \|\mathbf{U}^J\|_{\mathbf{A}^h}^2 + \|\mathbf{U}^j - \mathbf{U}^{j-1}\|_{2,h}^2\right] \leq \|\mathbf{U}^0\|_{2,h} + \sum_{j=1}^{J} \tau \|\mathbf{f}^j\|_{2,h}^2 \tag{2}$$

where

$$\|\mathbf{v}\|_{2,h}^2 = \sum_{i=1}^{N-1} h\mathbf{v}_i^2, \quad \|\mathbf{v}\|_{\mathbf{A}^h}^2 = h\mathbf{v}^T\mathbf{A}^h\mathbf{v}.$$

Remark 2.1 (discrete norms)**.** Note that we have the following Riemann sum approximation:

$$\|\mathbf{u}^j\|_{2,h}^2 = \sum_{i=1}^{N-1} hu(t_j, x_i)^2 \approx \int_0^1 u(t_j, x)^2 dx.$$

I claim that the $\mathbf{A}^h$ norm for a vector is a Riemann sum approximation of the square integral of a derivative

$$\|\mathbf{u}^j\|_{\mathbf{A}^h}^2 \approx \int_0^1 |u_x(t_j, x)|^2 dx.$$

To see this, we compute

$$\|\mathbf{u}^j\|_{\mathbf{A}^h}^2 = \sum_{i=1}^{N-1} hu(t_j, x_i) \left(\frac{-u(t_j, x_{i-1}) + 2u(t_j, x_i) - u(t_j, x_{i+1})}{h^2}\right)$$

We observe that this is a Riemann sum approximation for:

$$\|\mathbf{u}^j\|_{\mathbf{A}^h}^2 \approx -\int_0^1 u(t_j, x) u_{xx}(t_j, x) dx,$$

and we integrate by parts to get:

$$\|\mathbf{u}^j\|_{\mathbf{A}^h}^2 \approx -\int_0^1 u(t_j, x) u_{xx}(t_j, x) dx = \int_0^1 |u_x(t_j, x)|^2 dx$$

As a result, the estimate (2) can be viewed as a discrete analog to the energy estimates we have proved in earlier lectures:

$$\int_0^1 u(T, x)^2 dx + \int_0^T \int_0^1 u_x(t, x)^2 dx \leq \int_0^1 u(0, x)^2 dx + \int_0^T \int_0^1 f(t, x)^2 dx$$

The additional term:

$$\sum_{j=1}^{J} \left[\|\mathbf{U}^j - \mathbf{U}^{j-1}\|_{2,h}^2\right]$$

is known as **numerical dissipation**. This term shows that the numerical method dissipates more energy than expected from the PDE.

Proof of Discrete Energy Estimate. Just to mimic the continuous case, we take a dot product of the discrete evolution equation with $h\mathbf{U}^j$ to get

$$h\mathbf{U}^j \cdot D_\tau \mathbf{U}^j + \|\mathbf{U}^j\|^2_{\mathbf{A}^h} = h\mathbf{U}^j \cdot \mathbf{f}^j$$

We write out the first term

$$h\mathbf{U}^j \cdot D_\tau \mathbf{U}^j = \frac{h}{\tau}\mathbf{U}^j \cdot \left(\mathbf{U}^j - \mathbf{U}^{j-1}\right),$$

and use the following fact for vectors:

$$\mathbf{a} \cdot (\mathbf{a} - \mathbf{b}) = \frac{1}{2}\|\mathbf{a}\|_2^2 - \frac{1}{2}\|\mathbf{b}\|_2^2 + \frac{1}{2}\|\mathbf{a} - \mathbf{b}\|_2^2.$$

Hence, with $\mathbf{a} = \mathbf{U}^j$, and $\mathbf{b} = \mathbf{U}^{j-1}$, we have

$$\begin{aligned} h\mathbf{U}^j \cdot D_\tau \mathbf{U}^j &= \frac{h}{\tau}\left(\frac{1}{2}\|\mathbf{U}^j\|_2^2 - \frac{1}{2}\|\mathbf{U}^{j-1}\|_2^2 + \frac{1}{2}\|\mathbf{U}^j - \mathbf{U}^{j-1}\|_2^2\right) \\ &= \frac{1}{\tau}\left(\frac{1}{2}\|\mathbf{U}^j\|_{2,h}^2 - \frac{1}{2}\|\mathbf{U}^{j-1}\|_{2,h}^2 + \frac{1}{2}\|\mathbf{U}^j - \mathbf{U}^{j-1}\|_{2,h}^2\right). \end{aligned}$$

The discrete equation now reads

$$\frac{1}{\tau}\left(\frac{1}{2}\|\mathbf{U}^j\|_{2,h}^2 - \frac{1}{2}\|\mathbf{U}^{j-1}\|_{2,h}^2 + \frac{1}{2}\|\mathbf{U}^j - \mathbf{U}^{j-1}\|_{2,h}^2\right) + \|\mathbf{U}^j\|^2_{\mathbf{A}^h} = h\mathbf{U}^j \cdot \mathbf{f}^j$$

We estimate the RHS using Young's inequality

$$h\mathbf{U}^j \cdot \mathbf{f}^j = h\sum_{i=1}^{N-1} \mathbf{U}_i^j \mathbf{f}_i^j \le \frac{h}{2}\sum_{i=1}^{N-1}(\mathbf{U}_i^j)^2 + (\mathbf{f}_i^j)^2 = \frac{1}{2}\left(\|\mathbf{U}^j\|_{2,h}^2 + \|\mathbf{f}^j\|_{2,h}^2\right)$$

Just like in the proof of energy estimates for the heat equation, we need a Poincare inequality:

$$\|\mathbf{U}^j\|_{2,h}^2 \le \|\mathbf{U}^j\|^2_{\mathbf{A}^h},$$

which we will prove in the next Lemma. Once we have the above inequality, the discrete equation is now an inequality:

$$\frac{1}{\tau}\left(\frac{1}{2}\|\mathbf{U}^j\|_{2,h}^2 - \frac{1}{2}\|\mathbf{U}^{j-1}\|_{2,h}^2 + \frac{1}{2}\|\mathbf{U}^j - \mathbf{U}^{j-1}\|_{2,h}^2\right) + \|\mathbf{U}^j\|^2_{\mathbf{A}^h} \le \frac{1}{2}\|\mathbf{U}^j\|^2_{\mathbf{A}^h} + \frac{1}{2}\|\mathbf{f}^j\|_{2,h}^2$$

Subtracting $\frac{1}{2}\|\mathbf{U}^j\|^2_{\mathbf{A}^h}$ from both sides and multiplying both sides by 2 leads to

$$\frac{1}{\tau}\left(\|\mathbf{U}^j\|_{2,h}^2 - \|\mathbf{U}^{j-1}\|_{2,h}^2 + \|\mathbf{U}^j - \mathbf{U}^{j-1}\|_{2,h}^2\right) + \|\mathbf{U}^j\|^2_{\mathbf{A}^h} \le \|\mathbf{f}^j\|_{2,h}^2$$

Finally, multiplying by τ and summing from $j = 1, \ldots, J$ yields

$$\sum_{j=1}^{J}\left(\|\mathbf{U}^j\|_{2,h}^2 - \|\mathbf{U}^{j-1}\|_{2,h}^2\right) + \sum_{j=1}^{J}\tau\|\mathbf{U}^j\|^2_{\mathbf{A}^h} + \|\mathbf{U}^j - \mathbf{U}^{j-1}\|_{2,h}^2 \le \sum_{j=1}^{J}\tau\|\mathbf{f}^j\|_{2,h}^2$$

Notice the sum on the LHS telescopes: LHS sum telescopes

$$\sum_{j=1}^{J} \left(\|\mathbf{U}^j\|_{2,h}^2 - \|\mathbf{U}^{j-1}\|_{2,h}^2 \right) = \|\mathbf{U}^J\|_{2,h}^2 - \|\mathbf{U}^0\|_{2,h}^2,$$

which gives the result. □

The important Poincare inequality lemma we need is a property of the matrix $\mathbf{A}^h$.

Lemma 2.1 (eigenvalues of $\mathbf{A}^h$)**.** The eigenvalues of $\mathbf{A}^h \in \mathbb{R}^{(N-1)\times(N-1)}$ are

$$\lambda_k = \frac{4}{h^2} \sin^2\left(\frac{k\pi h}{2}\right), \quad k = 1, \dots, N-1$$

with eigenvectors

$$\mathbf{v}_i^k = \sin(k\pi x_i).$$

Proof. Left as exercise to verify that

$$\frac{1}{h^2}\left(-\mathbf{v}_{i-1}^k + 2\mathbf{v}_i^k - \mathbf{v}_{i-1}^k\right) = \lambda_k \mathbf{v}_i^k$$

□

A consequence of this property is that we have a discrete Poincare inequality:

Lemma 2.2 (discrete Poincare inequality)**.** For any vector $\mathbf{v}$, we have

$$\|\mathbf{v}\|_{2,h}^2 \le \lambda_1 \|\mathbf{v}\|_{2,h}^2 \le \|\mathbf{v}\|_{\mathbf{A}^h}^2 \le \lambda_{N-1} \|\mathbf{v}\|_{2,h}^2 \le \frac{4}{h^2} \|\mathbf{v}\|_{2,h}^2.$$

In particular, we have

$$\|\mathbf{v}\|_{2,h}^2 \le \|\mathbf{v}\|_{\mathbf{A}^h}^2$$

COMPUTATIONAL PDE LECTURE 15

LUCAS BOUCK

1. Outline of today

- Prove Error Estimate for Backward Euler
- Prove alternative stability estimates for Backward/Forward Euler

2. Finite Difference Methods for the Heat Equation

We are trying to solve:

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), \quad t \in (0,1), x \in (0,1) \\ u(t,0) = 0, u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

We now list with various time stepping schemes for the heat equation, which you implemented in recitation:

- **Forward Euler** (approximates differential equation at t_{j-1})

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^{j-1} = \mathbf{f}^{j-1}$$

- **Backward Euler** (approximates differential equation at t_j)

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^j = \mathbf{f}^j$$

2.1. Error estimate for Backward Euler.

Recall that we have shown consistency of both schemes.

Proposition 2.1 (consistency of schemes)**.** Let u solve (1). Then $\mathbf{u}_i^j = u(t_j, x_i)$ satisfies

- **Forward Euler**

$$D_\tau \mathbf{u}^j + \mathbf{A}^h \mathbf{u}^{j-1} = \mathbf{f}^{j-1} + \boldsymbol{\tau}^j_{\tau,h,fe}$$

- **Backward Euler**

$$D_\tau \mathbf{u}^j + \mathbf{A}^h \mathbf{u}^j = \mathbf{f}^j + \boldsymbol{\tau}^j_{\tau,h,be}$$

Date: October 02, 2023.

where there is a constant $C > 0$ such that the truncation errors satisfy:

$$\|\boldsymbol{\tau}^j_{\tau,h,fe}\|_\infty \leq C\left(\tau|u_{tt}|_{max} + h^2|u_{xxxx}|_{max}\right)$$
$$\|\boldsymbol{\tau}^j_{\tau,h,be}\|_\infty \leq C\left(\tau|u_{tt}|_{max} + h^2|u_{xxxx}|_{max}\right)$$

and

$$|u_{tt}|_{max} = \max_{t,x\in[0,1]} |u_{tt}(t,x)|, \quad |u_{ttt}|_{max} = \max_{t,x\in[0,1]} |u_{ttt}(t,x)|, \quad |u_{xxxx}|_{max} = \max_{t,x\in[0,1]} |u_{xxxx}(t,x)|$$

We also have shown stability of Backward Euler.

Proposition 2.2 (unconditional stability of Backward Euler)**.** Let $\mathbf{U}^j$ be the sequence of solutions to the Backward Euler scheme:

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^j = \mathbf{f}^j,$$

then the vector $\mathbf{U}^j$ satisfies the following discrete energy estimate:

$$\|\mathbf{U}^J\|^2_{2,h} + \sum_{j=1}^{J}\left[\tau\|\mathbf{U}^j\|^2_{\mathbf{A}^h} + \|\mathbf{U}^j - \mathbf{U}^{j-1}\|^2_{2,h}\right] \leq \|\mathbf{U}^0\|_{2,h} + \sum_{j=1}^{J}\tau\|\mathbf{f}^j\|^2_{2,h} \tag{2}$$

where

$$\|\mathbf{v}\|^2_{2,h} = \sum_{i=1}^{N-1} h\mathbf{v}_i^2, \quad \|\mathbf{v}\|^2_{\mathbf{A}^h} = h\mathbf{v}^T\mathbf{A}^h\mathbf{v}.$$

Recall that we have

$$\text{Stability} + \text{Consistency} \implies \text{Convergence}$$

We now combine the two results to show an error estimate.

Proposition 2.3 (error estimate of Backward Euler)**.** Let u solve (1). Let $\mathbf{U}^j$ be the sequence of solutions to the Backward Euler scheme:

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^j = \mathbf{f}^j$$

Then, we define $\mathbf{e}^j = \mathbf{u}^j - \mathbf{U}^j$. Let $T = J\tau$. The error satisfies for some $C > 0$:

$$\|\mathbf{e}^J\|^2_{2,h} + \sum_{j=1}^{J}\tau\|\mathbf{e}^j\|^2_{\mathbf{A}^h} \leq \|\mathbf{e}^0\|_{2,h} + \sum_{j=1}^{J}\tau\|\boldsymbol{\tau}^j_{\tau,h,be}\|^2_{2,h} \leq CT\left(\tau^2|u_{ttt}|^2_{max} + h^4|u_{xxxx}|^2_{max}\right) \tag{3}$$

Moreover,

$$\|\mathbf{e}^J\|_{2,h} \leq \sqrt{CT}\left(\tau|u_{ttt}|_{max} + h^2|u_{xxxx}|_{max}\right)$$

Proof. Recall that

$$D_\tau \mathbf{u}^j + \mathbf{A}^h \mathbf{u}^j = \mathbf{f}^j + \boldsymbol{\tau}^j_{\tau,h,be},$$

and

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^j = \mathbf{f}^j.$$

Subtracting these two equations leads to

$$D_\tau \mathbf{e}^j + \mathbf{A}^h \mathbf{e}^j = \boldsymbol{\tau}^j_{\tau,h,be}.$$

Applying the stability estimate (2) to $\mathbf{e}^j$ and the estimates on $\boldsymbol{\tau}^j_{\tau,h,be}$ give the desired result. □

2.2. **Alternative Stability Result.** Note that the stability result we used mimics the energy estimates of the heat equation. We can actually show an even better stability result by mimicking the maximum principle of the heat equation.

Proposition 2.4 (∞-norm stability of Euler)**.** Let $\mathbf{U}^j$ be the sequence of solutions to the Backward Euler scheme. Then the vector $\mathbf{U}^J$ satisfies:

$$\|\mathbf{U}^J\|_\infty \le \|\mathbf{U}^0\|_\infty + \sum_{j=1}^{J} \tau \|\mathbf{f}^j\|_\infty.$$

Let $\mathbf{U}^j$ be the sequence of solutions to the ~~Backward~~ Euler scheme. Then if

$$\tau \le \frac{h^2}{2} \tag{4}$$

the vector $\mathbf{U}^j$ satisfies:

$$\|\mathbf{U}^J\|_\infty \le \|\mathbf{U}^0\|_\infty + \sum_{j=0}^{J-1} \tau \|\mathbf{f}^j\|_\infty.$$

Proof of B.E. estimate. Suppose i is such that $\mathbf{U}^j_i = \|\mathbf{U}^j\|_\infty$. Then the Backward Euler iteration reads:

$$\text{LHS} = \left(1 + \frac{2\tau}{h^2}\right) \mathbf{U}^j_i - \frac{\tau}{h^2}\left(\mathbf{U}^j_{i-1} + \mathbf{U}^j_{i+1}\right) = \mathbf{U}^{j-1}_i + \tau \mathbf{f}^j_i \le \|\mathbf{U}^{j-1}\|_\infty + \tau \|\mathbf{f}^j_i\|_\infty.$$

The LHS can be estimated from below. This essentially uses the fact that $\mathbf{A}^h$ is an M matrix and $-\mathbf{U}^j_i = -\|\mathbf{U}^j\|_\infty \le -\mathbf{U}^j_{i\pm1}$:

$$\text{LHS} \ge \left(1 + \frac{2\tau}{h^2}\right) \mathbf{U}^j_i - \frac{2\tau}{h^2} \mathbf{U}^j_i = \mathbf{U}^j_i = \|\mathbf{U}^j\|_\infty.$$

We now have

$$\|\mathbf{U}^j\|_\infty \le \|\mathbf{U}^{j-1}\|_\infty + \tau \|\mathbf{f}^j_i\|_\infty.$$

If there is no i is such that $\mathbf{U}_i^j = \|\mathbf{U}^j\|_\infty$, we'd then have an i such that i is such that $-\mathbf{U}_i^j = \|\mathbf{U}^j\|_\infty$. Then repeat the argument with $-\mathbf{U}^j$. Then apply the inequality inductively to get the final result. □

Proof of F.E. estimate. We repeat the same argument as for B.E. Suppose $\mathbf{U}_i^j = \|\mathbf{U}^j\|_\infty$. The FE scheme reads

$$\mathbf{U}_i^j = \left(1 - \frac{2\tau}{h^2}\right)\mathbf{U}_i^{j-1} + \frac{\tau}{h^2}\left(\mathbf{U}_{i-1}^{j-1} + \mathbf{U}_{i+1}^{j-1}\right) + \tau \mathbf{f}_i^{j-1} = \text{ RHS.}$$

Since we assume

$$\tau \leq \frac{h^2}{2},$$

then

$$\left(1 - \frac{2\tau}{h^2}\right) \geq 0, \text{ and } \left(1 - \frac{2\tau}{h^2}\right)\mathbf{U}_i^{j-1} \leq \left(1 - \frac{2\tau}{h^2}\right)\|\mathbf{U}j - 1\|_\infty.$$

We can then estimate the RHS by

$$\text{RHS} \leq \left(1 - \frac{2\tau}{h^2}\right)\|\mathbf{U}^{j-1}\|_\infty + 2\frac{\tau}{h^2}\|\mathbf{U}^{j-1}\|_\infty + \tau\|\mathbf{f}_i^{j-1}\|_\infty = \|\mathbf{U}^{j-1}\|_\infty + \tau\|\mathbf{f}_i^{j-1}\|_\infty.$$

Hence,

$$\|\mathbf{U}^j\|_\infty \leq \|\mathbf{U}^{j-1}\|_\infty + \tau\|\mathbf{f}_i^{j-1}\|_\infty.$$

□

Remark 2.1 (CFL Condition)**.** For the stability of Forward Euler, we required

$$\tau \leq \frac{h^2}{2_{max}}. \tag{5}$$

This is known as a CFL (Courant-Friedrichs-Lewy) condition. Although Forward Euler is cheaper at each step than backward Euler, we pay with a restriction on the size of the time-step size.

COMPUTATIONAL PDE LECTURE 16

LUCAS BOUCK

1. Outline of today

- Prove ∞-norm estimates for finite difference methods for the heat equation
- Extend analysis to Crank-Nicolson

2. Finite Difference Methods for the Heat Equation

We are trying to solve:

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = f(t,x), \quad t \in (0,1), x \in (0,1) \\ u(t,0) = 0, u(t,1) = 0 \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

We now list with various time stepping schemes for the heat equation, which you implemented in recitation:

- **Forward Euler** (approximates differential equation at t_{j-1})

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^{j-1} = \mathbf{f}^{j-1}$$

- **Backward Euler** (approximates differential equation at t_j)

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \mathbf{U}^j = \mathbf{f}^j$$

2.1. Error estimate for Backward Euler.

Recall that we have shown consistency of both schemes.

Proposition 2.1 (consistency of schemes)**.** Let u solve (1). Then $\mathbf{u}_i^j = u(t_j, x_i)$ satisfies

- **Forward Euler**

$$D_\tau \mathbf{u}^j + \mathbf{A}^h \mathbf{u}^{j-1} = \mathbf{f}^{j-1} + \boldsymbol{\tau}^j_{\tau,h,fe}$$

- **Backward Euler**

$$D_\tau \mathbf{u}^j + \mathbf{A}^h \mathbf{u}^j = \mathbf{f}^j + \boldsymbol{\tau}^j_{\tau,h,be}$$

Date: October 06, 2023.

where there is a constant $C > 0$ such that the truncation errors satisfy:

$$\|\boldsymbol{\tau}^j_{\tau,h,fe}\|_\infty \leq C\left(\tau|u_{tt}|_{max} + h^2|u_{xxxx}|_{max}\right)$$
$$\|\boldsymbol{\tau}^j_{\tau,h,be}\|_\infty \leq C\left(\tau|u_{tt}|_{max} + h^2|u_{xxxx}|_{max}\right)$$

and

$$|u_{tt}|_{max} = \max_{t,x\in[0,1]} |u_{tt}(t,x)|, \quad |u_{ttt}|_{max} = \max_{t,x\in[0,1]} |u_{ttt}(t,x)|, \quad |u_{xxxx}|_{max} = \max_{t,x\in[0,1]} |u_{xxxx}(t,x)|$$

Recall that we have

Proposition 2.2 (∞-norm stability of Euler)**.** Let $\mathbf{U}^j$ be the sequence of solutions to the Backward Euler scheme. Then the vector $\mathbf{U}^J$ satisfies:

$$\|\mathbf{U}^J\|_\infty \leq \|\mathbf{U}^0\|_\infty + \sum_{j=1}^{J} \tau\|\mathbf{f}^j\|_\infty.$$

Let $\mathbf{U}^j$ be the sequence of solutions to the Backward Euler scheme. Then if

$$\tau \leq \frac{h^2}{2} \tag{2}$$

the vector $\mathbf{U}^j$ satisfies:

$$\|\mathbf{U}^J\|_\infty \leq \|\mathbf{U}^0\|_\infty + \sum_{j=0}^{J-1} \tau\|\mathbf{f}^j\|_\infty.$$

The above stability result leads to the following error estimate:

Proposition 2.3 (∞-norm error estimate of Euler)**.** Let $\mathbf{U}^j$ be the sequence of solutions to either the Forward or Backward Euler scheme. Then for any h, τ for B.E. or $\tau \leq \frac{h^2}{2}$ for F.E., the error vector $\mathbf{e}^J = \mathbf{u}^J - \mathbf{U}^J$ satisfies:

$$\|\mathbf{e}^J\|_\infty \leq CT\left(\tau|u_{tt}|_{max} + h^2|u_{xxxx}|_{max}\right).$$

Proof. We have that $\mathbf{e}^j = \mathbf{u}^j - \mathbf{U}^j$ solves

classic: consistency + stability ⇒ convergence

$$D_\tau \mathbf{e}^j + \mathbf{A}^h \mathbf{e}^j = \boldsymbol{\tau}^j_{\tau,h}$$

and apply stability result with estimates on truncation error. □

Remark 2.1 (stability influence on error estimates)**.** The kind of stability we prove determines what kind of error estimate we expect. The following table shows the relevant PDE stability, discrete stability, and error estimate.

PDE Stability	Energy Estimates	Max Principle
Discrete Analog	$\|\mathbf{U}^J\|^2_{2,h} \leq \|\mathbf{U}^0\|^2_{2,h} + \sum_{j=1}^J \tau\|\mathbf{f}^j\|_{2,h}$	$\|\mathbf{U}^J\|_\infty \leq \|\mathbf{U}^0\|_\infty + \sum_{j=1}^J \tau\|\mathbf{f}^j\|_{2,h}$
Error Estimate	$\|\mathbf{e}^J\|_{2,h} \leq \mathcal{O}(\sqrt{T}(h^2+\tau))$	$\|\mathbf{e}^J\|_\infty \leq \mathcal{O}(T(h^2+\tau))$

3. Extending to Crank-Nicolson

Recall the Crank Nicolson iteration is an approximation of the PDE at $t_{j-1/2}$:

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \frac{\mathbf{U}^j + \mathbf{U}^{j-1}}{2} = \mathbf{f}^{j-1/2}$$

and we had the following truncation error result.

Proposition 3.1 (consistency of CN)**.** Let u solve (1). Then $\mathbf{u}_i^j = u(t_j, x_i)$ satisfies

$$D_\tau \mathbf{u}^j + \frac{1}{2}\mathbf{A}^h(\mathbf{u}^j + \mathbf{u}^{j-1}) = \mathbf{f}^{j+1/2} + \boldsymbol{\tau}^j_{\tau,h,cn}$$

where there is a constant $C > 0$ such that the truncation error satisfies:

$$\|\boldsymbol{\tau}^j_{\tau,h,cn}\|_\infty \leq C\left(\tau^2|u_{ttt}|_{max} + \tau^2|u_{xxtt}|_{max} + h^2|u_{xxxx}|_{max}\right)$$

If we think of the relevant stability being energy estimates, then the discrete analog for Crank-Nicolson is

Proposition 3.2 (energy estimates of CN)**.** Let $\mathbf{U}^j$ solve the iteration:

$$D_\tau \mathbf{U}^j + \mathbf{A}^h \frac{\mathbf{U}^j + \mathbf{U}^{j-1}}{2} = \mathbf{f}^{j-1/2}.$$

always gonna multiply by h.

Then,

$$\|\mathbf{U}^J\|_{2,h} \leq \|\mathbf{U}^0\|_{2,h} + \sum_{j=1}^{J} \tau \|\mathbf{f}^{j-1/2}\|_{2,h}.$$

multiply by $h\left(\frac{U^j + U^{j-1}}{2}\right)$

Proof. Will be a HW exercise. Multiply equation with $\frac{\mathbf{U}^j + \mathbf{U}^{j-1}}{2}$ □

Once we have a stability estimate, we can then immediately know the correct error estimate.

Proposition 3.3 (error estimates of CN)**.** Let $\mathbf{U}^j$ solve the CN iteration: Then the error $\mathbf{e}^j = \mathbf{u}^j - \mathbf{U}^j$ satisfies

$$\|\mathbf{e}^J\|_{2,h} \leq TC\left(\tau^2|u_{ttt}|_{max} + \tau^2|u_{xxtt}|_{max} + h^2|u_{xxxx}|_{max}\right).$$

Proof. Write down CN iteration for $\mathbf{e}^j$ and apply stability with estimate on truncation error. □

understand the differences in proof techniques for 2, h stability (energy est.) and ∞ stability (maximum principle)

COMPUTATIONAL PDE LECTURE 17, 18, AND 19

LUCAS BOUCK

1. Outline of these lectures

- Cover the von Neumann stability theory of finite difference methods

2. Setup

We are trying to solve the heat equation on the whole real line with zero forcing:

$$\begin{cases} u_t(t,x) - u_{xx}(t,x) = 0, \quad t \in (0,1), x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

Notice that we no longer have boundary conditions since there is no boundary!

We now consider the following numerical set up

- meshsize $h > 0$
- time stepsize $\tau > 0$
- Uniformly spaced but infinite grid: $\{x_j\}_{j \in \mathbb{Z}}$ where $x_j = jh$
- Grid function $U^{h,\tau} : \{x_j\}_{j \in \mathbb{Z}} \to \mathbb{R}$. For simplicity, we denote

$$u_j^n = U^{h,\tau}(t_n, x_j)$$

Our familiar numerical schemes for the heat equation now read

- **Forward Euler**

$$D_\tau u_j^n - D_h^2 u_j^{n-1} = 0$$

- **Backward Euler**

$$D_\tau u_j^n - D_h^2 u_j^n = 0$$

- **Crank-Nicholson**

$$D_\tau u_j^n - D_h^2 \left(\frac{u_j^n + u_j^{n-1}}{2} \right) = 0$$

where

$$D_\tau u_j^n = \frac{u_j^n - u_j^{n-1}}{\tau}, \quad D_h^2 u_j^n = \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{h^2}$$

We now dig into the von Neumann theory.

Date: October 9-13, 2023.

3. Symbol of an Operator

We now introduce the concept of the symbol of an operator. Let $\xi \in \mathbb{R}$. We consider a special grid function

$$v_j = e^{i\xi x_j} = e^{i\xi hj},$$

where i is the imaginary unit. You may have seen Euler's Formula before:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta),$$

which will be helpful in our future calculations. Thus, we can interpret $v_j = e^{i\xi hj}$ as a grid function with oscillations like sine and cosine with frequency ξ.

Definition 3.1 (symbol)**.** Let A denote an linear map that maps grid functions to grid functions. We consider Av as an infinite matrix vector product:

$$(Au)_j = \sum_{k=-\infty}^{\infty} a_{jk}u_j$$

The **symbol** of the operator A is a function $S : \mathbb{R} \to \mathbb{C}$ such that for $\xi \in \mathbb{R}$ and $v_j = e^{i\xi hj}$, we have

$$S(\xi h)v_j = (Av)_j.$$

Remark 3.1 (interpretation of the symbol)**.** The symbol of an operator can be interpreted a few ways.

- $S(\xi h)$ is an eigenvalue of an operator A with eigenvector $v_j = e^{i\xi hj}$
- $S(\xi h)$ tells us how A acts on different frequencies ξ. For example, if $S(\xi h) = 2$, then an operator A will double the amplitude of $e^{i\xi hj}$.
- Recall that for a function defined by $f(x) = e^{i\xi x}$, we know

$$\frac{d}{dx}f(x) = if(x), \quad \frac{d^2}{dx^2}f(x) = -f(x),$$

so the second derivative of $e^{i\xi x}$ multiplies the function by -1. Essentially the symbol is an attempt to mimic these nice properties that we observe for derivatives at the discrete level.

3.1. Examples of symbols.

We now compute a few examples of symbols of operators.

- **Second finite difference** To compute the symbol of D_h^2, we write $S(\xi h)v_j = D_h^2 v_j$ and solve for $S(\xi h)$ with $v_j = e^{i\xi hj}$.

$$S(\xi h)v_j = D_h^2 v_j = \frac{v_{j+1} - 2v_j + v_{j-1}}{h^2} = \frac{e^{i\xi h(j+1)} - 2e^{i\xi hj} + e^{i\xi h(j-1)}}{h^2}$$

Recall the product property of exponentials: $e^{i\xi h(j\pm 1)} = e^{i\xi hj}e^{\pm i\xi h}$, so

$$S(\xi h)v_j = \frac{e^{i\xi hj}e^{i\xi h} - 2e^{i\xi hj} + e^{i\xi hj}e^{-i\xi h}}{h^2} = \frac{v_j}{h^2}\left(e^{i\xi h} + e^{-i\xi h} - 2\right).$$

We then use Euler's formula to simplify the complex exponentials

$$\begin{aligned} e^{i\xi h} + e^{-i\xi h} &= \cos(\xi h) + i\sin(\xi h) + \cos(-\xi h) + i\sin(-\xi h) \\ &= \cos(\xi h) + i\sin(\xi h) + \cos(\xi h) - i\sin(\xi h) \\ &= 2\cos(\xi h). \end{aligned}$$

Hence the symbol for D_h^2 is

$$S(\xi h) = \frac{2}{h^2}\left(\cos(\xi h) - 1\right).$$

- **Forward Euler Iteration** We consider the operator of A where $u_j^{n+1} = Au_j^n$ solves 1 iteration of Forward Euler. To compute the symbol of A, we write $u_j^n = v_j = e^{i\xi hj}$ and $u_j^{n+1} = S(\xi h)v_j$. The Forward Euler iteration now reads:

$$\frac{S(\xi h)v_j - v_j}{\tau} - D_h^2 v_j = 0$$

Rearranging and solving for $S(\xi h)v_j$ leads to

$$S(\xi h)v_j = v_j + \tau D_h^2 v_j$$

We now use the previous example $D_h^2 v_j = \frac{2}{h^2}\left(\cos(\xi h) - 1\right)v_j$ to simplify

$$S(\xi h)v_j = v_j + \frac{2\tau}{h^2}\left(\cos(\xi h) - 1\right)v_j$$

For simplicity of notation for the rest of the von Neumann stability lectures, I'll denote

$$\lambda = \frac{\tau}{h^2}.$$

With this notation, we have that the symbol of Forward Euler is

$$S(\xi h) = 1 + 2\lambda\left(\cos(\xi h) - 1\right)$$

3.2. **Instability of Forward Euler.** For this discussion, we consider a grid function that is as oscillatory as possible as in it switches between $+1$ and -1 at every grid point:

$$u_j^0 = (-1)^{|j|} = \cos(\pi j) = \cos(\pi j) + i\sin(\pi j) = e^{i\pi j},$$

which is the special grid function we we used to compute the symbol $S(\xi h)$ at $\xi h = \pi$. Hence, we may apply our previous work to see how Forward Euler acts on v_j. One step of Forward Euler is

$$u_j^1 = S(\xi h)u_j^0 = \left[1 + 2\lambda\left(\underbrace{\cos(\xi h)}_{=-1} - 1\right)\right] u_j^0 = (1 - 4\lambda)u_j^0.$$

Applying n steps of Forward Euler results in

$$u_j^1 = (1 - 4\lambda)^n u_j^0 = (1 - 4\lambda)^n (-1)^{|j|}.$$

The amplitude of these oscillations grow exponentially in n if $1-4\lambda < -1$ or $1-4\lambda > -1$. This instability occurs when

$$\lambda > \frac{1}{2}.$$

In terms of h, τ, this occurs when

$$\tau > \frac{h^2}{2}.$$

Recall from previous lectures that we required that $\tau \leq \frac{h^2}{2}$ (CFL condition) was a sufficient condition for stability of Forward Euler. This discussion shows that in fact, $\tau \leq \frac{h^2}{2}$ is also a necessary condition for stability.

3.3. **Examples of symbols of implicit schemes.** To find the symbols of Backward Euler and Crank-Nicholson, we apply the same procedure as Forward Euler by replacing $u_j^{n+1} = S(\xi h)u_j^n$ and $u_j^n = v_j e^{i\xi hj}$. We then solve for $S(\xi h)$. We now go over the symbols of these schemes.

- **Backward Euler** With the mentioned substitutions, the scheme reads:

$$\frac{S(\xi h)v_j - v_j}{\tau} - D_h^2[S(\xi h)v_j] = 0$$

Recall the symbol of $D_h^2 v_j = \frac{2}{h^2}\left(\cos(\xi h) - 1\right)$, so rearranging and multiplying by τ yields

$$(S(\xi h) - 1)v_j - S(\xi h)2\lambda\left(\cos(\xi h) - 1\right)v_j = 0.$$

where $\lambda = \tau/h^2$ as before. We then collect terms multiplying $S(\xi h)$ and add v_j to both sides to further simplify

$$S(\xi h)[1 - 2\lambda\left(\cos(\xi h) - 1\right)]v_j = v_j.$$

Hence, the symbol is

$$S(\xi h) = \frac{1}{1 + 2\lambda\left(1 - \cos(\xi h)\right)}.$$

- **Crank-Nicholson** With the mentioned substitutions, the scheme reads:

$$\frac{S(\xi h)v_j - v_j}{\tau} - D_h^2\left[\frac{S(\xi h)v_j + v_j}{2}\right] = 0.$$

I'll leave the computations as an exercise. The resulting symbol is

$$S(\xi h) = \frac{1-\lambda(1-\cos(\xi h))}{1+\lambda(1-\cos(\xi h))}.$$

4. General theory

We now state the general theory and leave the proofs for a later time. We first need a few definitions. For a grid function U, we define the following norm:

$$\|U\|_{2,h} = \left(h\sum_{j=-\infty}^{\infty}|u_j|^2\right)^{1/2}.$$

We have used finite sums of this form to study discrete energy estimates. Note that if $U_j = u(x_j)$ for some $u : \mathbb{R} \to \mathbb{R}$, then

$$\|U\|_{2,h} \approx \left(\int_{-\infty}^{\infty}|u(x)|^2 dx\right)^{1/2}$$

since the sum is a Riemann sum approximation of the integral.

Theorem 4.1 (stability)**.** Let A be an operator on grid functions. We have that A is stable:

$$\|AU\|_{2,h} \le \|U\|_{2,h}$$

if and only if the symbol of A satisfies the following bound for all $\xi \in \mathbb{R}$:

$$|S(\xi h)| \le 1$$

4.1. **Application of general theory to time stepping schemes.** We recall the symbols of the various time stepping schemes ($\lambda = \tau/h^2$):

- **Forward Euler:**

$$S(\xi h) = 1 + 2\lambda\left(\cos(\xi h) - 1\right)$$

- **Backward Euler:**

$$S(\xi h) = \frac{1}{1+2\lambda\left(1-\cos(\xi h)\right)}.$$

- **Crank-Nicholson:**

$$S(\xi h) = \frac{1-\lambda(1-\cos(\xi h))}{1+\lambda(1-\cos(\xi h))}$$

We now determine conditions on λ to guarantee stability of these schemes.

- **Forward Euler:** We want to determine λ so that for all ξ:
$$-1 \leq 1 + 2\lambda\left(\cos(\xi h) - 1\right) \leq 1,$$
which is equivalent to requiring
$$-2 \leq 2\lambda\left(\cos(\xi h) - 1\right) \leq 0.$$
Notice that $\cos(\xi h) - 1 \leq 0$, so all we need to check is
$$-2 \leq 2\lambda\left(\cos(\xi h) - 1\right).$$
The minimum of the RHS of the above inequality is achieved for $\xi h = \pi$, so we require $-2 \leq -4\lambda$, or
$$\tau \leq \frac{h^2}{2},$$
which is the CFL condition.
- **Backward Euler:** Notice that $0 \leq 1 - \cos(\xi h) \leq 2$ and
$$1 \leq 1 + 2\lambda\left(1 - \cos(\xi h)\right) \leq 1 + 4\lambda$$
Hence, the symbol of Backward Euler can be bounded above for all ξ:
$$|S(\xi h)| = \left|\frac{1}{1 + 2\lambda\left(1 - \cos(\xi h)\right)}\right| \leq 1,$$
and Backward Euler is unconditionally stable.
- **Crank-Nicholson:** It is left as an exercise to show that the symbol of Crank-Nicholson satisfies $|S(\xi h)| \leq 1$ for all ξ and for all choices of $\lambda \geq 0$. Hence, Crank-Nicholson is unconditionally stable.

5. Comparison of Methods in High Frequency Regime

In addition to a general theory, the von Neumann method allows us to look at the behavior of methods applied to different frequencies. We now look at the behavior of these methods in the presence of high frequencies.

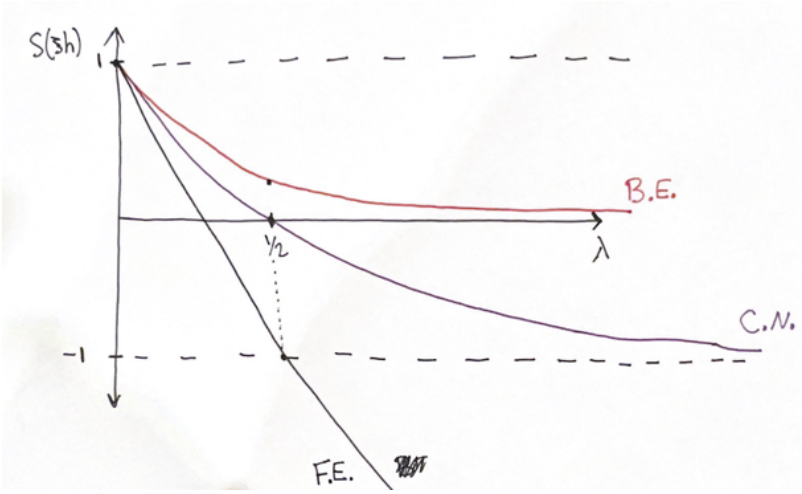
We consider a grid function that is as oscillatory as possible as in it switches between $+1$ and -1 at every grid point:
$$u_j = (-1)^{|j|} = \cos(\pi j) = \cos(\pi j) + i\sin(\pi j) = e^{i\pi j},$$
which is the special grid function we we used to compute the symbol $S(\xi h)$ at $\xi h = \pi$. The symbols of each method at $\xi h = \pi$ are

- **Forward Euler:**
$$S(\xi h) = 1 - 4\lambda$$
- **Backward Euler:**
$$S(\xi h) = \frac{1}{1 + 4\lambda}.$$

- **Crank-Nicholson:**

$$S(\xi h) = \frac{1 - 2\lambda}{1 + 2\lambda}$$

We now plot each symbol as a function of λ:



If we take τ to be relatively large compared with h^2, i.e. $\tau = h^\alpha$ for $0 \leq \alpha < 2$, then $\lambda \to \infty$ as $h \to 0$. The symbols then converge to

$$\lim_{\lambda \to \infty} S(\pi) = \begin{cases} -\infty, & \text{Forward Euler} \\ 0, & \text{Backward Euler} \\ -1, & \text{Crank-Nicholson} \end{cases}$$

These limits reveal disadvantages of each method.

- Forward Euler becomes unstable as $\lambda > 1/2$.
- Backward Euler over dissipates high frequencies as $\lambda \to \infty$
- Crank-Nicholson under dissipates high frequencies as $\lambda \to \infty$

5.1. **One method to rule them all: the θ-method.** An method that covers all the methods discussed here is the θ-method, which is

$$D_\tau u_j^n - D_h^2 \left(\theta u_j^n + (1 - \theta) u_j^{n-1} \right) = 0.$$

Note that the method reduces to FE if $\theta = 0$, B.E. if $\theta = 1$, and C.N. if $\theta = 1/2$. It is left as an exercise to check that the symbol of the θ-method is

$$S(\xi h) = \frac{1 + 2\lambda(\theta - 1)(1 - \cos(\xi h))}{1 + 2\lambda\theta(1 - \cos(\xi h))}.$$

note the θ method

If we look at the high frequency behavior of the θ-method, we take $\xi h = \pi$ and compute

$$S(\pi) = \frac{1 + 4\lambda(\theta - 1)}{1 + 4\lambda\theta}.$$

whose limit as $\lambda \to \infty$ is

$$\lim_{\lambda \to \infty} S(\pi) = \begin{cases} \frac{(\theta-1)}{\theta}, & \theta > 0 \\ -\infty, & \theta = 0 \end{cases}.$$

The θ parameter gives us more flexibility in tuning our method for different situations. Note that the limit is $\lim_{\lambda\to\infty} S(\pi) < -1$ if $\theta < 1/2$. As a result, we expect a CFL condition for $\theta < 1/2$. Also, one can prove the method is unconditionally stable for $\theta \geq 1/2$.

COMPUTATIONAL PDE LECTURES 20 AND 21

LUCAS BOUCK

1. Outline of these lectures

- Derivation of transport equation
- Method of characteristics

2. Setup

We are trying to solve the transport equation on the whole real line:

$$\begin{cases} u_t(t,x) + cu_x(t,x) = 0, & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

where $c \neq 0$ is some constant speed.

3. Derivation of the Transport Equation

Suppose that u represents a mass density with units [kg/m] of some particles blowing in the wind where the wind velocity is c. A physical property of this system is **conservation of mass**, i.e. for

$$M(t) = \int_{ct}^{b+ct} u(t,x)dx$$

we have

$$\frac{d}{dt}M(t) = 0.$$

We can rewrite conservation of mass to derive the transport equation (1). Computing the time derivative, we use Leibniz integral rule:

$$\begin{aligned} \frac{d}{dt}M(t) &= \int_{ct}^{b+ct} u_t(t,x)dx + u(b+ct,x)\frac{d}{dt}(b+ct) - u(ct,x)\frac{d}{dt}(ct) \\ &= \int_{ct}^{b+ct} u_t(t,x)dx + c\big(u(b+ct,x) - u(ct,x)\big) \end{aligned}$$

Date: October 23-25, 2023.

For these derivations, we typically want to compare apples to apples, i.e. we want everything as an integral over the same domain. Hence, we use Fundamental Theorem of Calculus to write:

$$u(b+ct,x) - u(ct,x) = \int_{ct}^{b+ct} u_x(t,x)dx,$$

which results in

$$0 = \int_{ct}^{b+ct} u_t(t,x) + cu_x(t,x)dx.$$

Dividing the above equation by b and taking a limit as $b \to$ shows that

$$0 = \lim_{b\to 0} \frac{1}{b} \int_{ct}^{b+ct} u_t(t,x) + cu_x(t,x)dx. = u_t(ct,x) + cu_x(ct,x).$$

Notice that ct is arbitrary so we have for any t, x:

$$u_t(t,x) + cu_x(t,x) = 0,$$

which is the desired PDE (1).

Remark 3.1 (other conservation laws)**.** The transport equation bears resemblance to many conservation laws of the form:

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

where f is the flux of some quantity u that is to be conserved. By applying Fundamental Theorem of Calculus to the RHS of the above equation and repeating arguments we used for the transport equation, we have

$$u_t(t,x) + \partial_x\big(f(u(t,x))\big) = 0,$$

which is referred to as a **nonlinear conservation law**. One such famous example is Burger's equation:

$$u_t + \partial_x\left(\frac{1}{2}u^2\right) = 0,$$

which is a common equation in fluid dynamics.

4. Solving the transport equation: method of characteristics

We are interested in solving (1):

$$\begin{cases} u_t(t,x) + cu_x(t,x) = 0, & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases}.$$

We first begin with a geometric idea. Note that the RHS of (1) reads

$$0 = u_t(t,x) + cu_x(t,x) = \begin{pmatrix} 1 \\ c \end{pmatrix} \cdot \begin{pmatrix} u_t \\ u_x \end{pmatrix}.$$

Notice that the directional derivative of u in the direction $(1, c)$ is zero. Hence u must be constant along a line of the form $\{(t,x) : x = x_0 + ct\}$. This line is known as a **characteristic**.

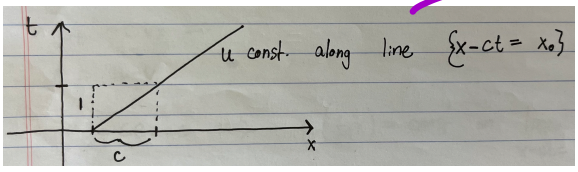


FIGURE 1. Characteristic lines where u is constant along the line.

Importantly, we then have that for any t:

$$u(t, x_0 + ct) = u(0, x_0) = u_0(x_0).$$

To then get the solution formula for u, we write $x_0 = x - ct$ to get

$$u(t,x) = u_0(x - ct). \tag{2}$$

We summarize this technique in the following proposition.

Proposition 4.1 (method of characteristics solution)**.** Let $u_0 \in C^1(\mathbb{R})$. Then u defined in (2) solves (1).

Proof. Clearly $u(0,x) = u_0(x)$. To check the PDE, we use chain rule:

$$\partial_t u(t,x) + c\partial_x u(t,x) = \partial_t[u_0(x-ct)] + c\partial_x[u_0(x-ct)] = -cu_0'(x-ct) + cu_0'(x-ct) = 0.$$

□

Remark 4.1 (physical interpretation of characteristics)**.** If a particle starts at x_0, then the line $\{(t,x) : x_0 + ct = x\}$ is the physical path that a particle takes in space.

Method of characteristics also shows us uniqueness.

Proposition 4.2 (method of characteristics solution)**.** Let $u_0 \in C^1(\mathbb{R})$. The solution to (1) is unique.

Proof. Let u, v be solutions to (1). Then their difference $w = u - v$ solves

$$\begin{cases} w_t(t,x) + cw_x(t,x) = 0, & t > 0, x \in \mathbb{R} \\ w(0,x) = 0 \end{cases}.$$

Recall that we know that w is constant along lines $x_0 + ct$. Hence,

$$w(t,x) = w(0, x - ct) = 0,$$

which completes the proof. □

Since we have an explicit solution formula and know that the solution is unique, we also have stability.

Proposition 4.3 (stability of transport)**.** Let $u_0 \in C^1(\mathbb{R})$ such that $\max_{z\in\mathbb{R}} |u_0(z)|$ is well-defined. The solution u to (1) satisfies for all t, x:

$$|u(t,x)| \leq \max_{z\in\mathbb{R}} |u_0(z)|$$

Proof. We use the solution formula in (2): $|u(t,x)| = |u_0(x-ct)| \leq \max_{z\in\mathbb{R}} |u_0(z)|$. □

4.1. **Method of characteristics with forcing.** We now consider a small generalization of (1) to

$$\begin{cases} u_t(t,x) + cu_x(t,x) = f(t,x), & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{3}$$

The motivation for this addition is that when we solve with finite differences, we introduce a small truncation error that will behave like f. To solve (3), we again look at the solution along characteristics. We define

$$v(t) = u(t, x_0 + ct),$$

which is u evaluated along a characteristic. Computing the time derivative of v with chain rule yields

$$v'(t) = u_t(t, x_0 + ct) + cu_x(t, x_0 + ct) = f(t, x_0 + ct),$$

and evaluating v at 0 gives

$$v(0) = u(0, x_0) = u_0(x_0).$$

To solve this ODE initial value problem, we only need to apply Fundamental Theorem of Calculus:

$$v(t) = v(0) + \int_0^t v'(s)ds = u_0(x_0) + \int_0^t f(s, x_0 + cs)ds.$$

We then use $x = x_0 + ct$ or $x - ct = x_0$ to get

$$u(t,x) = v(t) = u_0(x - ct) + \int_0^t f(s, x_0 + c(s-t))ds. \tag{4}$$

which solves (3). One can apply the arguments from the previous section to say that (3) admits a unique solution. Hence, we can use the solution (4) to state the following stability result.

Proposition 4.4 (stability of transport with forcing)**.** Let $u_0 \in C^1(\mathbb{R})$ such that $\max_{z\in\mathbb{R}} |u_0(z)|$ is well-defined. Let $f \in C^1((0,\infty)\times\mathbb{R})$ be such that $\max_{s>0,z\in\mathbb{R}} |f(s,z)|$ is well-defined. Let u be a solution to (3). Then u satisfies

$$|u(t,x)| \leq \max_{z\in\mathbb{R}} |u_0(z)| + t \max_{s>0,z\in\mathbb{R}} |f(s,z)|$$

Proof. We can again use an explicit solution formula (4):

$$u(t,x) = u_0(x-ct) + \int_0^t f(s, x_0 + c(s-t))ds.$$

We then take the absolute value and use triangle inequality

$$\begin{aligned} |u(t,x)| &\leq |u_0(x-ct)| + \left| \int_0^t f(s, x_0 + c(s-t))ds \right| \\ &\leq \max_{z\in\mathbb{R}} |u_0(z)| + \left| \int_0^t f(s, x_0 + c(s-t))ds \right| \\ &\leq \max_{z\in\mathbb{R}} |u_0(z)| + \int_0^t \left| f(s, x_0 + c(s-t)) \right| ds \\ &\leq \max_{z\in\mathbb{R}} |u_0(z)| + t \max_{s>0,z\in\mathbb{R}} |f(s,z)| \end{aligned}$$

which is the desired result. □

Note that this stability formula will be more useful in the analysis of a numerical method since the numerical method will have a small f, that is a truncation error. Our finite difference scheme will try to mimic this property.

4.2. **Method of characteristics for more general situations.** Lots of things in the worlds do not have constant velocity, so it makes sense to model systems where the velocity of a particle at a point is $c(t,x)$ and depends on space and time. In this case, we can following the physical derivation at the beginning of these notes to find a new transport equation:

$$u_t(t,x) + c(t,x)u_x(t,x) + c_x(t,x)u(t,x) = 0.$$

This particular equation now looks quite different from the transport equation with constant velocity. However, the method of characteristics still works as long as we remember that we an easily solve for u along the path a particle takes.

4.2.1. *General procedure.* We consider the following equation

$$\begin{cases} u_t(t,x) + c(t,x)u_x(t,x) + b(t,x)u(t,x) = f(t,x), \quad t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{5}$$

We break the solution into two steps:
Step 1: Find the characteristics We first search for the characteristics of (5). With the interpretation that this the path a particle takes with velocity $c(t,x)$, we find a solution to the ODE:

$$\begin{cases} y'(t) = c(t, y(t)), t > 0 \\ y(0) = x_0 \end{cases}.$$

Solve ODE for characteristics

The solution y describes the position of a particle that started at position x_0 and evolved with velocity $c(t, y(t))$.

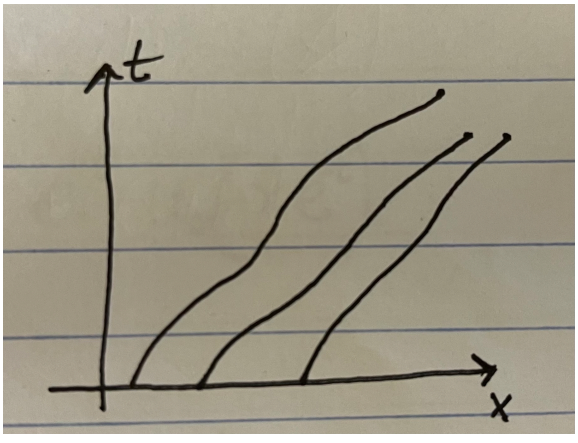


FIGURE 2. Characteristics $\{(t, y(t)) : t > 0\}$ that solve $y'(t) = c(t, y(t))$

Step 2: Solve for u along characteristics Similar to the case with forcing, we solve for u along characteristics. That is, we define $v(t) = u(t, y(t))$, and find the ODE for v. We have from chain rule and the PDE:

$$v'(t) = u_t(t, y(t)) + u_x(t, y(t))y'(t) = -b(t, y(t))u(t, y(t)) + f(t, y(t)) = -b(t, y(t))v(t) + f(t, y(t)).$$

We also have the initial condition

$$v(0) = u(0, x_0) = u_0(x_0).$$

Combining these leads to solving the initial value problem:

Solve ODE to find $v(t)$

$$\begin{cases} v'(t) = -b(t, y(t))v(t) + f(t, y(t)), t > 0 \\ v(0) = u_0(x_0) \end{cases}.$$

We complete the construction by writing x_0 in terms of $x = y(t)$ and t.

complete by writing x_0 in terms of $x = y(t)$ and t

4.2.2. *Examples.* We now look at 2 examples.

Example 4.1. Consider $c(t, x) = x, f(t, x) = 0, b(t, x) = 1$, so

$$\begin{cases} u_t(t, x) + xu_x(t, x) + u = 0, \quad t > 0, x \in \mathbb{R} \\ u(0, x) = u_0(x) \end{cases}$$

The ODE for characteristic in Step 1 is

$$\begin{cases} y'(t) = y(t), t > 0 \\ y(0) = x_0 \end{cases}$$

whose solution is $y(t) = x_0 e^t$.

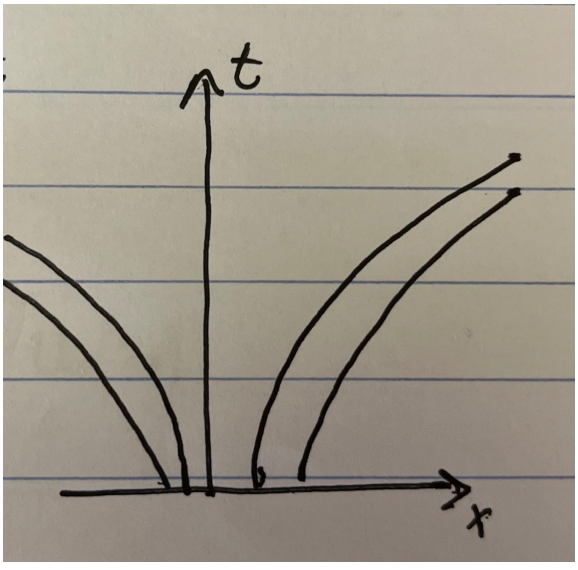


FIGURE 3. Characteristics $y(t) = x_0 e^t$

The ODE for the solution $u(t, y(t)) = v(t)$ along a characteristic in Step 2 is

$$\begin{cases} v'(t) = -v(t), t > 0 \\ v(0) = u(x_0) \end{cases}$$

so

$$u(t, y(t)) = v(t) = e^{-t} u_0(x_0).$$

Setting $x = y(t)$, we have $x_0 = xe^{-t}$, and the solution is

$$u(t, x) = e^{-t} u_0(xe^{-t}).$$

Example 4.2. Consider $c(t, x) = t, b(t, x) = 0$, and $f(t, x) = 1$. The transport equation reads:

$$\begin{cases} u_t(t, x) + tu_x(t, x) = 1, \quad t > 0, x \in \mathbb{R} \\ u(0, x) = u_0(x) \end{cases}$$

The ODE for characteristic in Step 1 is

$$\begin{cases} y'(t) = t, t > 0 \\ y(0) = x_0 \end{cases}$$

whose solution is $y(t) = x_0 + \frac{t^2}{2}$.

The ODE for the solution $u(t, y(t)) = v(t)$ along a characteristic in Step 2 is

$$\begin{cases} v'(t) = 1, t > 0 \\ v(0) = u(x_0) \end{cases}$$

so

$$u(t, y(t)) = v(t) = 1 + u_0(x_0).$$

Writing $x = y(t) = x_0 + \frac{t^2}{2}$, we have the solution

$$u(t, x) = 1 + u_0\left(x - \frac{t^2}{2}\right)$$

COMPUTATIONAL PDE LECTURE 22

LUCAS BOUCK

1. Outline of this lecture

- Start numerics for transport equation

2. Setup

We are trying to solve the transport equation on the whole real line:

$$\begin{cases} u_t(t,x) + cu_x(t,x) = f(t,x), & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

where $c \neq 0$ is some constant speed. We consider the following numerical setup similar to our study of von Neumann stability analysis.

- meshsize $h > 0$
- time stepsize $\tau > 0$
- Uniformly spaced but infinite grid: $\{x_j\}_{j\in\mathbb{Z}}$ where $x_j = jh$
- Grid function $U^{h,\tau} : \{x_j\}_{j\in\mathbb{Z}} \to \mathbb{R}$. For simplicity, we denote

$$\mathbf{U}_j^n = U^{h,\tau}(t_n, x_j)$$

Also, recall the discrete time derivative

$$D_\tau \mathbf{U}_j^n = \frac{\mathbf{U}_j^n - \mathbf{U}_j^{n-1}}{\tau}.$$

Our goals for designing a numerical method again follow the Lax postulate, which states:

$$\text{stability} + \text{consistency} \implies \text{convergence}$$

We will easily have consistency due to Taylor expansions. Our main focus will be to design a stable method. In particular, we want the method to satisfy

$$\|\mathbf{U}^n\|_\infty \leq \|\mathbf{U}^{n-1}\|_\infty + \tau\|\mathbf{f}^{n-1}\|_\infty,$$

which mimics the stability we saw for the transport equation at the continuous level.

Date: October 27, 2023.

upwind method :
forward euler with backward
finite diff in space

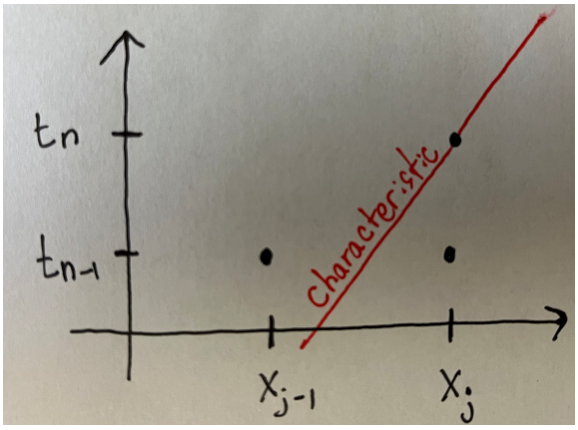


FIGURE 1. Red is the true characteristic and the black dots are the numerical stencil. We can see that the stencil "looks upwind".

3. First method: upwind scheme

We will assume in this section that $c > 0$, meaning the wind is going left to right. The upwind scheme is a Forward Euler method with a backward finite difference in space:

$$D_\tau \mathbf{U}_j^n + c\left(\frac{\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}}{h}\right) = \mathbf{f}_j^{n-1}. \tag{2}$$

FE with backward fd in space

Essentially, this scheme "looks upwind" to compute the solution to the transport equation. The figure below shows the upwind scheme

The truncation error of this scheme will be first order, due to the use of backward differences.

Proposition 3.1 (consistency of upwind scheme)**.** Let $\mathbf{u}_j^n = u(t_n, x_j)$, where u is an exact solution of (1). Then $\mathbf{u}_j^n$ solves the discrete equation

$$D_\tau \mathbf{u}_j^n + c\left(\frac{\mathbf{u}_j^{n-1} - \mathbf{u}_{j-1}^{n-1}}{h}\right) = \mathbf{f}^{n-1} + \boldsymbol{\tau}_{h,\tau}^{n-1}, \tag{3}$$

backward fd's :
$\tau |u_{tt}|_{max}$
$h |u_{xx}|_{max}$

where $\boldsymbol{\tau}_{h,\tau}^{n-1}$ is a truncation error and satisfies

$$\|\boldsymbol{\tau}_{h,\tau}^{n-1}\|_\infty \leq C\left(h|u_{xx}|_{max} + \tau|u_{tt}|_{max}\right)$$

Proof. Use Taylor expansion at (t_{n-1}, x_j) similar to Forward Euler for the heat equation. □

taylor expand at (t_{n-1}, x_j)

upwind iteration :

$$D_\tau U_j^n + c\left(\frac{U_j^{n-1} - U_{j-1}^{n-1}}{h}\right) = f_j^{n-1}$$

$$\hookrightarrow \frac{U_j^n - U_j^{n-1}}{\tau} + c\left(\frac{U_j^{n-1} - U_{j-1}^{n-1}}{h}\right) = f_j^{n-1} \quad : \quad U_j^n - U_j^{n-1} + \frac{c\tau}{h}\left(U_j^{n-1} - U_{j-1}^{n-1}\right) = \tau f_j^{n-1}$$

A more delicate issue is stability of the upwind scheme, which holds under a CFL condition.

Proposition 3.2 (stability of upwind scheme)**.** Let $\mathbf{U}_j^n$ be the discrete solution of the upwind scheme. Further assume that the following CFL condition holds

$$\tau \leq \frac{h}{c}.$$

Then,

$$\|\mathbf{U}^n\|_\infty \leq \|\mathbf{U}^{n-1}\|_\infty + \tau \|\mathbf{f}^{n-1}\|_\infty.$$

Proof. The proof follows similarly to that of stability of Forward Euler for the heat equation under a CFL condition.

We first suppose that we have j such that $\mathbf{U}_j^n = \max_{k\in\mathbb{Z}} |\mathbf{U}_k^n| = \|\mathbf{U}^n\|_\infty$. If there is no such positive $\mathbf{U}_j^n$, then we look at $-\mathbf{U}_j^n$. We now write the upwind iteration:

$$\mathbf{U}_j^n = \left(1 - \frac{c\tau}{h}\right) \mathbf{U}_j^{n-1} + \frac{c\tau}{h} \mathbf{U}_{j-1}^{n-1} + \tau \mathbf{f}_j^{n-1}$$

Since we are "upwinding", we can see that $\frac{c\tau}{h} \geq 0$, so

$$\frac{c\tau}{h} \mathbf{U}_{j-1}^{n-1} \leq \frac{c\tau}{h} \|\mathbf{U}^{n-1}\|_\infty.$$

Also, the CFL condition tells us $\left(1 - \frac{c\tau}{h}\right) \geq 0$, so

$$\left(1 - \frac{c\tau}{h}\right) \mathbf{U}_j^{n-1} \leq \left(1 - \frac{c\tau}{h}\right) \|\mathbf{U}^{n-1}\|_\infty.$$

We combine these estimates to get

$$\|\mathbf{U}^n\|_\infty = \mathbf{U}_j^n \leq \|\mathbf{U}^{n-1}\|_\infty + \tau \|\mathbf{f}^{n-1}\|_\infty,$$

which is the desired result. □

Note that we can also sum the stability from $n = 1, \ldots, N$ to get

$$\|\mathbf{U}^n\|_\infty \leq \|\mathbf{U}^0\|_\infty + \tau \sum_{n=1}^{N} \|\mathbf{f}^{n-1}\|_\infty,$$

which is a more useful stability result for the error estimate below.

Proposition 3.3 (error estimate of upwind scheme)**.** Let $\mathbf{U}_j^n$ be the discrete solution of the upwind scheme. Further assume that the following CFL condition holds

$$\tau \leq \frac{h}{c}.$$

Then, if $\mathbf{u}_j^n = u(t_n, x_j)$, we have the error estimate:

$$\|\mathbf{u}^n - \mathbf{U}^n\|_\infty \leq t_n C \left(h|u_{xx}|_{max} + \tau |u_{tt}|_{max}\right).$$

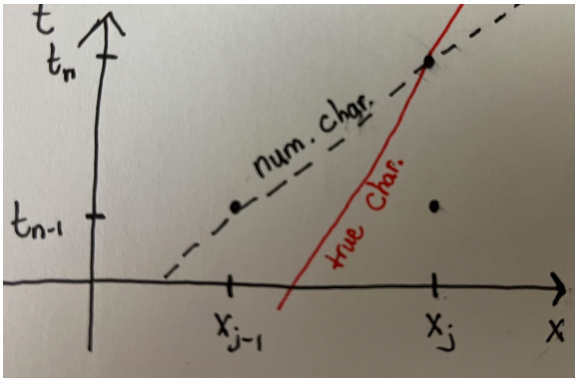


FIGURE 2. Red is the true characteristic and the black dashed line is the "numerical characteristic." The CFL condition states essentially that the numerical characteristic needs to move faster than the true characteristic.

Proof. Use typical techniques we learned from heat equation. Write down a discrete equation for the error $\mathbf{u}^n - \mathbf{U}^n$. Combine the stability estimate and the estimate on truncation error to complete the result. □

Remark 3.1 (CFL condition for transport)**.** Note that the proof required the following CFL condition:

$$\tau \leq \frac{h}{c}.$$

I claim that this is a mild condition and is the best we can do for transport problems. The reason is that the fundamental property of transport equations are characteristics. Given a point (t_j, x_i), the numerical method needs to use enough points at t_{j-1} to capture the true characteristic. Note that the characteristic line (t_j, x_i) goes through t_{j-1} at $x = x_i - \tau c$. Thus, in order for $x \in (x_{i-1}, x_{i+1})$, we require that $\tau \leq \frac{h}{c}$.

One benefit of knowing that the CFL condition $\tau \leq \frac{h}{c}$ is the best we can expect means we can get away with a Forward Euler (hence cheaper to compute) method for transport. For the heat equation, we had a tradeoff between Backward Euler (expensive to compute) and Forward Euler (cheap to compute but bad CFL condition).

COMPUTATIONAL PDE LECTURE 23

LUCAS BOUCK

1. Outline of this lecture

- Continue with discussion of upwinding

2. Setup

We are trying to solve the transport equation on the whole real line:

$$\begin{cases} u_t(t,x) + cu_x(t,x) = f(t,x), & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

We will assume in this section that $c > 0$, meaning the wind is going left to right. The upwind scheme is a Forward Euler method with a backward finite difference in space:

$$D_\tau \mathbf{U}_j^n + c\left(\frac{\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}}{h}\right) = \mathbf{f}_j^{n-1}.$$

Essentially, this scheme "looks upwind" to compute the solution to the transport equation.

3. Why upwinding "works"

There are two reasons why the above technique works. One is that upwinding introduces some numerical diffusion, and the second is that upwinding respects conservation of mass.

3.1. Numerical Diffusion.

We again look at the upwinding scheme:

$$D_\tau \mathbf{U}_j^n + c\left(\frac{\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}}{h}\right) = \mathbf{f}^{n-1},$$

and observe that

$$\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1} = \frac{1}{2}(\mathbf{U}_{j+1}^{n-1} - \mathbf{U}_{j-1}^{n-1}) + \frac{1}{2}\left(-\mathbf{U}_{j+1}^{n-1} + 2\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}\right).$$

Date: October 30, 2023.

Rearranging the upwind scheme leads to

$$D_\tau \mathbf{U}_j^n + \frac{c}{2h}(\mathbf{U}_{j+1}^{n-1} - \mathbf{U}_{j-1}^{n-1}) - \frac{h}{2h^2}\left(-\mathbf{U}_{j+1}^{n-1} + 2\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}\right) = \mathbf{f}^{n-1}.$$

Note that

$$\frac{1}{2h}(\mathbf{U}_{j+1}^{n-1} - \mathbf{U}_{j-1}^{n-1})$$

is a centered finite difference approximation of $u_x(t_{n-1}, x_j)$, and

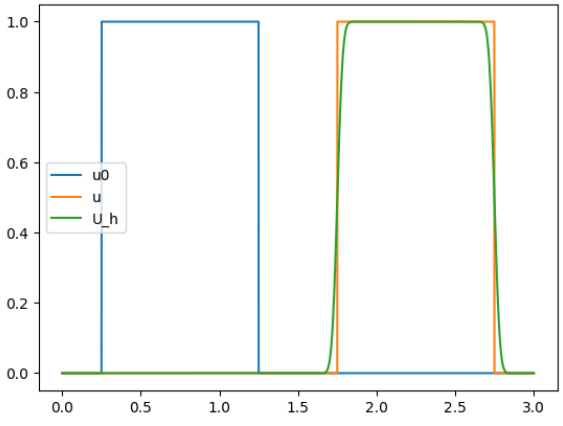
$$\frac{-\mathbf{U}_{j+1}^{n-1} + 2\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}}{h^2}$$

is our typical finite difference approximation for $-u_{xx}(t_n, x_j)$. Hence, the upwind scheme is numerically solving a modified heat equation:

$$u_t + cu_x - \frac{ch}{2}u_{xx} = f.$$

where the coefficient of diffusion goes to zero with $h \to 0$. From recitation, you can see that with a rough initial condition, the upwinding method adds some smoothing, the discrete solution looks similar to the true solution, but is smoother. All the methods we study will introduce some numerical diffusion.

FIGURE 1. Initial condition in blue. Exact solution in orange and numerical solution in green. Notice that the numerical method introduces some smoothing of the solution.



3.2. **Conservation of mass at numerical level.** Recall that the transport equation with $f = 0$ has a conservation of mass property:

$$\frac{d}{dt}\int_a^b u(t,x)dx = -cu(t,b) + cu(t,a).$$

where cu was a mass flux. The upwind scheme also enjoys a similar property at the discrete level if $f = 0$.

Let

$$M_h(t_n) = h\sum_{j=a}^{b} \mathbf{U}_j^n.$$

upwind mimics conservation of mass

⇓

Scheme is conservative

Then,

$$D_\tau M_h(t_n) = h\left(\sum_{j=a}^{b} D_\tau \mathbf{U}_j^n\right)$$

We then use the upwind scheme iteration $D_\tau \mathbf{U}_j^n = -c\frac{\mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}}{h}$ to further expand

$$D_\tau M_h(t_n) = -c\left(\sum_{j=a}^{b} \mathbf{U}_j^{n-1} - \mathbf{U}_{j-1}^{n-1}\right) = -c\mathbf{U}_b^{n-1} + c\mathbf{U}_{a-1}^{n-1}$$

Thus, the upwind scheme mimics the conservation of mass property at the discrete level. Schemes that do this are known as **conservative**.

Remark 3.1 (upwind for nonlinear conservation laws)**.** In fact, upwinding can be a good scheme for nonlinear conservation laws of the form.

$$u_t + \partial_x(f(u)) = 0.$$

Recall the above equation has the following conservation property:

* Linear conserv. laws on steroids

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

Assuming $f'(u) \geq 0$ for all u, the following upwind scheme is a reasonable choice for a nonlinear conservation law:

$$D_\tau \mathbf{U}_j^n + \left(\frac{f(\mathbf{U}_j^{n-1}) - f(\mathbf{U}_{j-1}^{n-1})}{h}\right) = 0$$

because it is also conservative in the sense that

$$D_\tau\left(h\sum_{j=a}^{b} D_\tau \mathbf{U}_j^n\right) = -f(\mathbf{U}_b^{n-1}) + f(\mathbf{U}_a^{n-1}),$$

so the upwind scheme outlined here is a reasonable choice for Burger's equation $f(u) = \frac{u^2}{2}$ as long as $u_0 \geq 0$.

4. Other methods for transport

Upwinding is fine, but lower order. We now discuss some strategies to get higher order methods.

4.1. Centered Difference is unstable.

A natural next scheme to achieve higher accuracy in space would be to replace the upwind term with a centered finite difference. The resulting iteration is

$$D_\tau \mathbf{U}_j^n + c\left(\frac{\mathbf{U}_{j+1}^{n-1} - \mathbf{U}_{j-1}^{n-1}}{2h}\right) = \mathbf{f}_j^{n-1}. \tag{2}$$

However, this scheme is always unstable! The reason is that a backward or forward difference introduces numerical diffusion (think of backward Euler for the heat equation), while a centered difference does not introduce extra dissipation (think of Crank Nicholson for the heat equation). We can prove this statement more directly with von Neumann analysis.

Proposition 4.1 (instability of centered difference scheme)**.** Suppose $\mathbf{f}_j^{n-1} = 0$. The symbol of the centered difference scheme in (2) is

$$S(\xi h) = 1 - \frac{ic\tau}{h}\sin(\xi h).$$

In particular, we have

$$|S(\pi/2)| > 1,$$

so the centered difference scheme in (2) is always unstable.

Proof. The computation is an exercise. The procedure is We compute the symbol by replacing $\mathbf{U}_j^{n-1} = \mathbf{v}_j = e^{i\xi hj}$ and $\mathbf{U}_j^n = S(\xi h)\mathbf{v}_j$, and then we isolate the symbol. The key step you need is the space derivative term

$$\begin{aligned}\mathbf{v}_{j+1} - \mathbf{v}_{j-1} &= e^{i\xi h(j+1)} - e^{i\xi h(j-1)} = e^{i\xi hj}\left(e^{i\xi h} - e^{-i\xi h}\right)\\ &= \mathbf{v}_j\left(\cos(\xi h) + i\sin(\xi h) - \cos(\xi h) + i\sin(\xi h)\right)\\ &= 2i\sin(\xi h)\mathbf{v}_j.\end{aligned}$$

□

COMPUTATIONAL PDE LECTURE 24

LUCAS BOUCK

1. Outline of this lecture

- Discuss Lax-Fredrich's scheme and derive the Lax-Wendroff scheme

2. Setup

We are trying to solve the transport equation on the whole real line:

$$\begin{cases} u_t(t,x) + cu_x(t,x) = f(t,x), & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

3. Other methods for transport

Recall that upwind is stable, but lower order. We now discuss some strategies to get higher order methods.

3.1. **Centered Difference is unstable.** A natural next scheme to achieve higher accuracy in space would be to replace the upwind term with a centered finite difference. The resulting iteration is

$$D_\tau \mathbf{U}_j^n + c\left(\frac{\mathbf{U}_{j+1}^{n-1} - \mathbf{U}_{j-1}^{n-1}}{2h}\right) = \mathbf{f}_j^{n-1}. \tag{2}$$

However, this scheme is always unstable! We proved this using von Neumann Stability analysis

3.2. **Another scheme: Lax-Frederich's Scheme.** In order to save the centered difference, we now look at a way to "add diffusion" or "smooth out" the numerical solution. One way to smooth out is to either add a diffusion term directly or by taking an average. The Frederich's scheme works by taking an average :

$$\frac{1}{\tau}\left[\mathbf{U}_j^n - \left(\frac{\mathbf{U}_{j-1}^{n-1} + \mathbf{U}_{j+1}^{n-1}}{2}\right)\right] + c\left(\frac{\mathbf{U}_{j+1}^{n-1} - \mathbf{U}_{j-1}^{n-1}}{2h}\right) = \mathbf{f}_j^{n-1}. \tag{3}$$

to save a centered difference we need to add diffusion

⇓

lax-fredericks does this

save centered diff by adding numerical dissipation

Date: November 01, 2023.

One can show that this scheme is stable and is a second order scheme (in h for fixed τ) for the following modified PDE:

$$u_t + cu_x - \frac{h^2}{2\tau}u_{xx} = 0.$$

However, note that we want to take $\tau \leq \frac{h}{c}$, so if $\tau = \frac{h}{c}$ and $h \to 0$, we no longer have a second order scheme due to the $\frac{h^2}{2\tau} \approx h$ numerical diffusion term.

We now state the specific results for the Lax-Frederich's scheme in (3) but leave the proofs as exercises.

Proposition 3.1 (stability of Lax-Frederich's)**.** Let $\mathbf{U}^n$ solve the Frederich's iteration in (3). Assume the CFL condition $\tau \leq \frac{h}{c}$. Then,

$$\|\mathbf{U}^n\|_\infty \leq \|\mathbf{U}^{n-1}\|_\infty + \tau\|\mathbf{f}^{n-1}\|_\infty.$$

Proposition 3.2 (consistency of Lax-Frederich's)**.** Let $\mathbf{u}_j^n = u(t_n, x_j)$ be the exact solution of the transport equation. Then,

$$\frac{1}{\tau}\left[\mathbf{u}_j^n - \left(\frac{\mathbf{u}_{j-1}^{n-1} + \mathbf{u}_{j+1}^{n-1}}{2}\right)\right] + c\left(\frac{\mathbf{u}_{j+1}^n - \mathbf{u}_{j-1}^n}{2h}\right) = \mathbf{f}_j^{n-1} + \boldsymbol{\tau}_j^{n-1}.$$

where there is a $C > 0$ independent of h, τ such that

$$|\boldsymbol{\tau}_j^{n-1}| \leq C\left(h^2|u_{xxx}|_{max} + \tau|u_{tt}|_{max} + \frac{h^2}{\tau}|u_{xx}|_{max}\right)$$

Proposition 3.3 (error estimate of Lax-Frederich's)**.** Let $\mathbf{U}^n$ solve the Frederich's iteration in (3). Assume the CFL condition $\tau \leq \frac{h}{c}$. Let $\mathbf{u}_j^n = u(t_n, x_j)$ be the exact solution of the transport equation. Then, the error satisfies

$$\|\mathbf{u}^n - \mathbf{U}^n\|_\infty \leq Ct_n\left(h^2|u_{xxx}|_{max} + \tau|u_{tt}|_{max} + \frac{h^2}{\tau}|u_{xx}|_{max}\right).$$

3.3. **Second order method: Lax-Wendroff.** We now derive a second order scheme using the symbol of an operator. Suppose $u(t, x) = e^{i\xi x}$. Then if u solves the transport equation, we have

$$u(t + \tau, x) = e^{i\xi(x - c\tau)} = e^{-i\xi c\tau}u(t, x).$$

In some sense, the true symbol we'd like to approximate is $S(\xi) = e^{-i\xi c\tau}$. If we are able to approximate the true symbol at $\xi = 0$ for ξ small, then we are likely to have a convergent method. One way to do this would be to write a Taylor expansion of the symbol

$$S(\xi) = 1 - c\tau i\xi - \frac{c^2\tau^2}{2}\xi^2 + \mathcal{O}(\xi^3).$$

This can be summarized in the following theorem that we will not prove.

Theorem 3.1. We define an explicit finite difference method by the following iteration:

$$\mathbf{U}_j^{n+1} = \sum_{k=-\infty}^{\infty} a_k \mathbf{U}_{j+k}^n.$$

Let $S(\xi h)$ be the symbol of the above iteration. That is, if $\mathbf{v}_j = e^{i\xi h j}$, then

$$S(\xi h)\mathbf{v}_j = \sum_{k=-\infty}^{\infty} a_k \mathbf{v}_{j+k}.$$

Further assume that for for $\tau \leq \frac{h}{c}$, we have $|S(\xi h)| \leq 1$ for all ξ and there is a $\xi^* > 0$ such that if $|\xi h| < \xi^*$, we have

$$|S(\xi h) - e^{-i\xi c\tau}| \leq C|\xi h|^{r+1}.$$

Then the finite difference method satisfies the following error estimate for $\tau \leq \frac{h}{c}$:

$$\|\mathbf{u}^n - \mathbf{U}^n\|_{2,h} \leq \mathcal{O}(t_n(\tau^r + h^r)).$$

3.3.1. *Derivation of Lax-Wendroff.* We won't prove this theorem, but will use it to derive a second order scheme. In order to achieve $|S(\xi) - e^{-i\xi c\tau}| \leq C|\xi|^3$, we need to match the Taylor expansion of the true symbol. We define our scheme with three coefficients

$$\mathbf{U}_j^{n+1} = a_{-1}\mathbf{U}_{j-1}^n + a_0\mathbf{U}_j^n + a_1\mathbf{U}_{j+1}^n,$$

and now compute the symbol as

$$S(\xi h) = (a_{-1}e^{-i\xi h} + a_0 + a_1 e^{i\xi h}),$$

with

$$S'(\xi h) = (-iha_{-1}e^{-i\xi h} + iha_1 e^{i\xi h})$$

and

$$S''(\xi h) = (-h^2 a_{-1}e^{-i\xi h} - h^2 a_1 e^{i\xi h})$$

We now want

$$S(0) = 1, \quad S'(0) = -c\tau i, \quad S''(0) = -\frac{c^2\tau^2}{2}$$

Setting $S(0) = 1$ leads to $a_{-1} + a_0 + a_1 = 0$. We compute

$$S'(0) = -iha_{-1} + a_1 hi.$$

Hence, we want $-\frac{c\tau}{h} = a_1 - a_{-1}$. Finally, we compute

$$S''(0) = -h^2 a_{-1} - h^2 a_1,$$

and require $a_{-1} + a_1 = \frac{c^2\tau^2}{2h^2}$. We are left with the linear system with $\lambda = \tau/h$:

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{-1} \\ a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 1 \\ -c\lambda \\ c^2\lambda^2 \end{pmatrix}$$

whose solution is

$$a_{-1} = \frac{1}{2}(c^2\lambda^2 + c\lambda), \quad a_0 = 1 - c^2\lambda^2, \quad a_1 = \frac{1}{2}(c^2\lambda^2 - c\lambda).$$

We now have a second order scheme, which is the Lax-Wendroff scheme:

$$\mathbf{U}_j^{n+1} = \frac{1}{2}(c^2\lambda^2 + c\lambda)\mathbf{U}_{j-1}^n + (1 - c^2\lambda^2)\mathbf{U}_j^n + \frac{1}{2}(c^2\lambda^2 - c\lambda)\mathbf{U}_{j+1}^n$$

COMPUTATIONAL PDE LECTURE 25

LUCAS BOUCK

1. Outline of this lecture

- Discuss schemes for nonlinear conservation laws

2. Setup

We are trying to solve the following nonlinear conservation law on the whole real line:

$$\begin{cases} u_t(t,x) + \partial_x[f(u(t,x)] = 0, \quad t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x) \end{cases} \tag{1}$$

An important concept that solutions to (2) is mass conservation

$$\frac{d}{dt}\int_{\mathbb{R}} u(t,x)dx = 0,$$

which we have shown in previous lectures. An important consequence of mass conservation is what is known as order preservation.

Proposition 2.1 (order preservation)**.** Let u, v be two solutions (2) with $u(0,x) = u_0(x)$ and $v(0,x) = v_0(x)$. If $u_0(x) \leq v_0(x)$ for all $x \in \mathbb{R}$, then $u(t,x) \leq v(t,x)$ for all $x \in \mathbb{R}$, $t > 0$.

Proof. A fact that we first use about is that the difference of two solutions is controlled by the difference of the initial conditions in an integral sense, i.e.

$$\int_{\mathbb{R}} |v(t,x) - u(t,x)|dx \leq \int_{\mathbb{R}} |v_0(x) - u_0(x)|dx$$

This fact is actually quite difficult to prove. Once we have it though, we use the assumption that $v_0(x) - u_0(x) \geq 0$ and we can drop the absolute value and apply the mass conservation property

$$\int_{\mathbb{R}} |v_0(x) - u_0(x)|dx = \int_{\mathbb{R}} v_0(t,x) - u_0(t,x)dx = \int_{\mathbb{R}} v(t,x) - u(t,x)dx.$$

Date: November 03, 2023.

Hence, we have

$$\int_{\mathbb{R}} |v(t,x) - u(t,x)|dx \leq \int_{\mathbb{R}} v(t,x) - u(t,x)dx.$$

This implies $v(t,x) \geq u(t,x)$ because if not, then the integral on the LHS would become larger and contradict the above inequality. □

3. Order preserving numerical schemes

A successful numerical method for (2) will mimic the order preservation property. This property is known as monotonicity:

Definition 3.1 (monotonicity)**.** We say a numerical iteration that maps $\mathbf{U}^n$ to $\mathbf{U}^{n+1}$ is monotone if for $\mathbf{V}_j^n \geq \mathbf{U}_j^n$ for all j, then $\mathbf{V}_j^{n+1} \geq \mathbf{U}_j^{n+1}$.

We now list some monotone schemes for (2).

- **Upwind** Suppose $f'(u) \geq 0$ for all $u \in \mathbb{R}$, then the upwind scheme is

$$\frac{\mathbf{U}_j^{n+1} - \mathbf{U}_j^n}{\tau} + \frac{f(\mathbf{U}_j^n) - f(\mathbf{U}_{j-1}^n)}{h} = 0$$

- **Lax-Fredrichs** The Lax-Fredrichs scheme is

$$\frac{\mathbf{U}_j^{n+1} - \frac{1}{2}(\mathbf{U}_{j+1}^n + \mathbf{U}_{j-1}^n)}{\tau} + \frac{f(\mathbf{U}_{j+1}^n) - f(\mathbf{U}_{j-1}^n)}{2h} = 0$$

These schemes are both monotone under assumptions on f. We first list the result for upwind:

Proposition 3.1 (monotonicity of upwind)**.** Suppose $f'(u) \geq 0$ for all $u\mathbb{R}$ and $f''(u) \geq 0$ for all $u \in \mathbb{R}$, i.e. f is convex. Then upwind is monotone as long as the CFL condition $\tau \leq \frac{h}{|f'(u_0)|_{max}}$.

Proof. We only proof the result for $f(u) = cu$ for $c > 0$, which is the transport equation. Assume $\mathbf{V}_j^n \geq \mathbf{U}_j^n$ for all j. We now write

$$\begin{aligned} \mathbf{U}_j^{n+1} &= \mathbf{U}_j^n + \frac{\tau c}{h}(\mathbf{U}_{j-1}^n - \mathbf{U}_j^n) \\ \mathbf{V}_j^{n+1} &= \mathbf{V}_j^n + \frac{\tau c}{h}(\mathbf{V}_{j-1}^n - \mathbf{V}_j^n) \end{aligned}$$

We now subtract to get

$$\mathbf{V}_j^{n+1} - \mathbf{U}_j^{n+1} = \left(1 - \frac{\tau c}{h}\right)\left(\mathbf{V}_j^n - \mathbf{U}_j^n\right) + \frac{\tau c}{h}\left(\mathbf{V}_{j-1}^n - \mathbf{U}_{j-1}^n\right).$$

Notice that by the CFL condition, $1 - \frac{\tau c}{h} \geq 0$, and by the fact that we used upwinding, we have $\frac{\tau c}{h} \geq 0$. Since we also have $\mathbf{V}_j^n - \mathbf{U}_j^n \geq 0$ for all j, then $\mathbf{V}_j^{n+1} \geq \mathbf{U}_j^{n+1}$ for all j, which is the desired result □

Remark 3.1. We note that the proof of monotonicity for the transport equation is essentially the proof of ∞ norm stability of upwind. In fact, monotonicity implies ∞ norm stability.

Remark 3.2. To prove the above monotonicity result for generic f that is convex, you want to repeat the same arguments but use properties of convex functions like

$$f(y) \geq f(x) + f'(x)(y - x),$$

which means the tangent line of a convex function lies below the graph of said function.

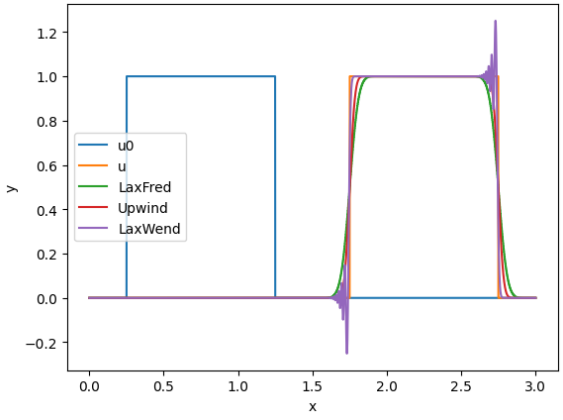
We also have that Lax-Fredrichs is monotone, though we will will not prove it.

Proposition 3.2 (monotonicity of Lax-Fredrichs)**.** Suppose $f''(u) \geq 0$ for all $u \in \mathbb{R}$, i.e. f is convex. Then Lax-Fredrichs is monotone as long as the CFL condition $\tau \leq \frac{h}{|f'(u_0)|_{max}}$.

Remark 3.3 (Lax-Wendroff is not monotone)**.** The Lax-Wendroff scheme we developed in previous lectures is not monotone for the transport equation. This can be seen by looking at the following initial condition from Recitation

$$u_0(x) = \begin{cases} 1, & x \in (1/4, 5/4), \\ 0, & \text{otherwise} \end{cases}.$$

We can see numerically in the next plot that Lax-Wendroff is not monotone, while upwind and Lax-Fredrichs are monotone.



COMPUTATIONAL PDE LECTURES 26-31

LUCAS BOUCK

1. Outline of these lectures

- Derive wave equation
- Solve the wave equation on the real line using a technique due to d'Alembert.
- Show energy estimates for the wave equation on the real line
- Discuss separation of variables for the wave equation.
- Explicit finite difference method for the wave equation: consistency and von Neumann analysis
- Derivation of implicit but energy conserving finite difference method for the wave equation

2. Setup

We will first look at the wave equation on the whole real line:

$$\begin{cases} u_{tt}(t,x) - c^2 u_{xx}(t,x) = 0, & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x), & x \in \mathbb{R} \\ u_t(0,x) = g_0(x), & x \in \mathbb{R} \end{cases} \tag{1}$$

Note that we now have a new initial condition, which specifies the initial velocity $u_t(0,x) = g_0(x)$. This is because we now have two derivatives in time on u.

3. Derivation of wave equation

We first begin with a derivation of the wave equation (1). Here, u represents the height of a string with mass density ρ that has constant tension, T, throughout the string. We now look at a free-body diagram of the string

Date: November 06-20, 2023.

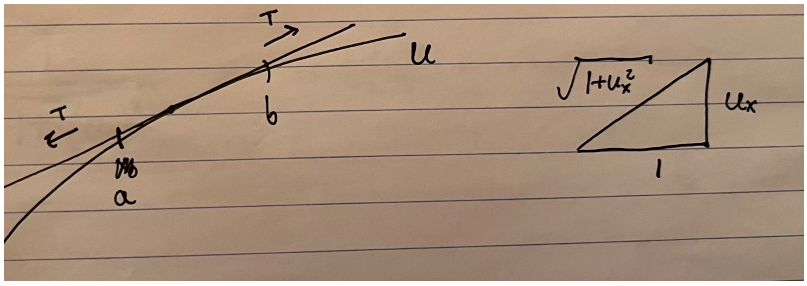


FIGURE 1. Free body diagram of string on left. The triangle on the right shows the decomposition of the forces into vertical and horizontal components.

Writing out Newton's second law $F = ma$, where F is the force, m is the mass, and a is the acceleration, we now break the forces into a vertical and horizontal component. The horizontal component is

$$\frac{T}{\sqrt{1+u_x(t,b)^2}} - \frac{T}{\sqrt{1+u_x(t,a)^2}} = 0.$$

The vertical component is

$$\frac{Tu_x(t,b)}{\sqrt{1+u_x(t,b)^2}} - \frac{Tu_x(t,a)}{\sqrt{1+u_x(t,a)^2}} = \int_a^b \rho u_{tt}(t,x)dx.$$

Note that the vertical component equation is of the form $F = ma$, since we have the vertical component of the tension force on the LHS, and the RHS is an integral of density times acceleration, which has units of mass times acceleration after an integral in space.

In order to compare apples to apples, we now rewrite the LHS in terms of an integral using Fundamental Theorem of Calculus and we have

$$\int_a^b \partial_x \left[\frac{Tu_x(t,x)}{\sqrt{1+u_x(t,x)^2}}\right] dx = \int_a^b \rho u_{tt}(t,x)dx.$$

We now use the same usual trick by dividing by $b-a$ and taking a limit as $b \to a$ to get

$$\rho u_{tt}(t,x) - \partial_x \left[\frac{Tu_x(t,x)}{\sqrt{1+u_x(t,x)^2}}\right] = 0$$

This equation is highly nonlinear and is difficult to study. To get the wave equation, we now make an assumption that u_x is small or $u_x \approx 0$. Then we have the Taylor

expansion:

$$\frac{1}{\sqrt{1+u_x(t,x)^2}} = 1 - \frac{u_x(t,x)^2}{2} + \mathcal{O}(u_x^4),$$

so

$$\frac{Tu_x(t,x)}{\sqrt{1+u_x(t,x)^2}} = Tu_x(t,x) + \mathcal{O}(u_x^2).$$

smallness assumption : $u_x \approx 0$ used in physical derivation

We then apply the derivative in ∂_x to get

$$\partial_x \left[\frac{Tu_x(t,x)}{\sqrt{1+u_x(t,x)^2}} \right] \approx Tu_{xx}(t,x).$$

The resulting equation after making this smallness assumption is the wave equation in (1)

$$u_{tt}(t,x) - c^2 u_{xx}(t,x) = 0,$$

where $c = \sqrt{\frac{T}{\rho}}$. Note that c here has units of speed. We'll see later that waves travel at speed c.

Remark 3.1 (smallness assumption)**.** Recall that we had to make a small amplitude assumption, i.e. $u_x \approx 0$. This is a common modeling technique to go from a nonlinear equation to a linear equation. Often, many linear theories are approximations of the more correct nonlinear theory, where we assume some quantity is small. For instance, if we have a steady state solution of the nonlinear wave equation ($u_{tt} = 0$), then we have the equation

$$-\partial_x \left[\frac{u_x(t,x)}{\sqrt{1+u_x(t,x)^2}} \right] = 0,$$

which is the equation of a line minimizing arc length. The linearized theory (assuming $u_x \approx 0$) is

$$-u_{xx}(t,x) = 0,$$

which is Laplace's equation.

Another example if you ever take a course in solid mechanics is you'll most likely study linear elasticity. This theory is making assumptions that the deformation of the material is small. The richer set of theories that are nonlinear are able to describe large deformations better.

4. d'Alembert's solution of the wave equation

We now want to construct a solution to (1). The main idea due to d'Alembert was to factor the equation into:

$$0 = u_{tt}(t,x) - c^2 u_{xx}(t,x) = (\partial_{tt} - c^2\partial_{xx})u(t,x) = (\partial_t - c\partial_x)(\partial_t + c\partial_x)u(t,x).$$

construct solution : factor eqn

$0 = u_{tt} - c^2 u_{xx} = (\partial_t - c^2\partial_{xx})u = (\partial_t - c\partial_x)(\partial_t + c\partial_x)u$

$u_t + cu_x = w$

Then w solves transport $w_t - cw_x = 0$

This factorization is motivated by $(a+b)(a-b) = a^2 - b^2$ for any two real numbers a, b. Hence, if we define $(\partial_t + c\partial_x)u(t,x) = w(t,x)$, then w solves a transport equation $(\partial_t - c\partial_x)w(t,x) = 0$. We then have the following system of transport equations:

$$\begin{cases} u_t + cu_x = w \\ w_t - cw_x = 0 \end{cases}$$

resulting system of transport equations

We know how to solve transport equations using the method of characteristics. We have that u is

$$u(t,x) = u_0(x-ct) + \int_0^t w(s, x + c(s-t))ds,$$

from equation 1 in the system

and w is

$$w(t,x) = w_0(x+ct) = \partial_t u(0, x+ct) - c\partial_x u(0, x+ct) = g_0(x+ct) - cu_0'(x+ct).$$

We now substitute the solution for w into the integral in the solution formula for u in order to write everything in terms of the initial data. We have

$$w(s, x+c(s-t)) = w_0(x+c(s-t)+cs) = g_0(x+c(s-t)+cs) + cu_0'(x+c(s-t)+cs),$$

so

write out in terms of U transport formula

$$u(t,x) = u_0(x-ct) + \int_0^t g_0(x+c(s-t)+cs) + cu_0'(x+c(s-t)+cs)ds.$$

The integral terms we can simplify using a change of variables $y = x + c(s-t) + cs$, so then

change of var for integral

$$u(t,x) = u_0(x-ct) + \frac{1}{2c}\int_{x-ct}^{x+ct} g_0(y) + cu_0'(y)dy.$$

The last term can be further simplified using Fundamental Theorem of Calculus

$$\frac{1}{2}\int_{x-ct}^{x+ct} u_0'(y)dy = \frac{1}{2}\left(u_0(x+ct) - u_0(x-ct)\right).$$

Hence,

$$u(t,x) = \frac{1}{2}\left[u_0(x+ct) + u_0(x-ct)\right] + \frac{1}{2c}\int_{x-ct}^{x+ct} g_0(y)dy, \tag{2}$$

which is the d'Alembert solution to the wave equation (1). This technique shows two important properties of the wave equation.

Example 4.1 (plucked string)**.** We now look at a simple example, where

$$u_0(x) = \begin{cases} 1 - |x|, & |x| \le 1 \\ 0, & \text{otherwise} \end{cases}$$

d'lambert solution :

$u = \frac{1}{2}[u_0(x+ct) + u_0(x-ct)] + \frac{1}{2c}\int_{x-ct}^{x+ct} g_0(y)dy$

(x^*, t^*)

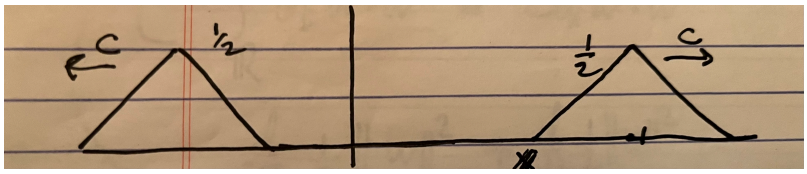
$x^* + ct^*$ $x^* - ct^*$

need information in $[x^* - ct^*, x^* + ct^*]$

and $g_0(x) = 0$. This is if we pull a string up with height 1. The solution is then

$$u(t,x) = \frac{1-|x+ct|}{2} + \frac{1-|x-ct|}{2},$$

which are two triangular waves of height $\frac{1}{2}$ traveling with velocities $-c$ and $+c$.



two triangular waves with speed −c and +c

FIGURE 2. Solution to wave equation of plucked string.

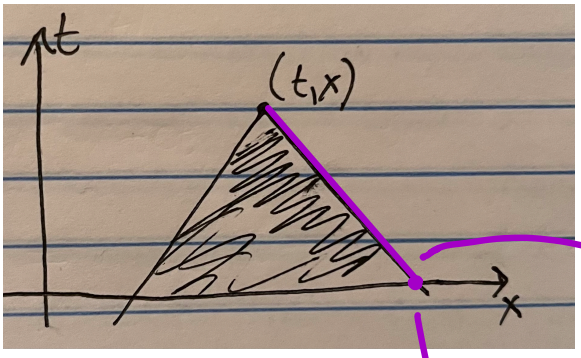
Remark 4.1 (domain of dependence)**.** Let (t^*, x^*) be a point in space-time. We then have that $u(t^*, x^*)$ depends on the initial values $(t = 0)$ in the interval $[t^* - cx^*, t^* + cx^*]$. More generally, we can modify the d'Alembert formula in (2) for any $t \leq t^*$:

$$u(t^*, x^*) = \frac{1}{2}\left[u(t^*, x^* + c(t^* - t)) + u_0(t^*, x^* - c(t^* - t))\right] + \frac{1}{2c}\int_{x^*-c(t^*-t)}^{x^*+c(t^*-t)} u_t(t,x)dx.$$

This change in the formula shows that the solution $u(t^*, x^*)$ also depends on information of u on the set $\{t\} \times [x^* - c(t^* - t), x^* + c(t^* - t)]$. Looking at all $0 \leq t \leq t^*$, we have that $u(t^*, x^*)$ depends on u in the cone

$$C_{t^*,x^*} = \{(t,x) : x^* - c(t^* - t) \leq x \leq x^* + c(t^* - t)\}.$$

change in formula
now u depends on set
$\{t\} \times [x^* - c(t^*-t),\ x^* + c(t^*-t)]$
clearly cone shaped
derive this information dependency from the d'Alembert solution



$u(t^*, x^*)$ depends on information about u in the cone
$x^* - c(t^*-t) \leq x \leq x^* + c(t^*-t)$
$x^* + c(t^*-t)$
$x^* + ct^*$

FIGURE 3. Domain of dependence of wave equation in shaded region.

This cone is known as the **domain of dependence**. We can also see that information travels at speed c. This will suggest a CFL condition $\tau \leq h/c$ for the numerics.

Remark 4.2 (coupled transport equations)**.** The factoring of the operator $(\partial_{tt} - c^2\partial_{xx}) = (\partial_t - c\partial_x)(\partial_t + c\partial_x)$ led to us solving a system of coupled transport equations:

$$\begin{cases} u_t + cu_x = w \\ w_t - cw_x = 0 \end{cases}$$

The above set of equations suggests that we can might be able to use our techniques from the transport equation to solve the wave equation. We'll also use the above structure to derive energy estimates and show uniqueness of solutions to the wave equation.

5. Energy estimates for the wave equation and uniqueness

This section derives energy estimates for the wave equation and will consequently show uniqueness. In order to have a result that will translate to numerics more easily. We consider the wave equation with forcing.

$$\begin{cases} u_{tt}(t,x) - c^2 u_{xx}(t,x) = f(t,x), & t > 0, x \in \mathbb{R} \\ u(0,x) = u_0(x), & x \in \mathbb{R} \\ u_t(0,x) = g_0(x), & x \in \mathbb{R} \end{cases} \tag{3}$$

We now derive energy estimates for the wave equation.

Proposition 5.1 (first energy estimate)**.** Let u be a C^2 solution to (3) with fast decay at $\pm\infty$. Then,

$$\frac{1}{2}\int_{\mathbb{R}} u(t,x)^2 + (u_t(t,x) + cu_x(t,x))^2 dx$$
$$\leq e^{2t}\left(\frac{1}{2}\int_{\mathbb{R}} u(0,x)^2 + (u_t(0,x) + cu_x(0,x))^2 dx + \int_0^t \int_{\mathbb{R}} f(s,x)^2 dx ds\right).$$

Proof. We write $w = u_t + cu_x$ and write the system of coupled transport equations:

$$\begin{aligned} u_t + cu_x &= w \\ w_t - cw_x &= f. \end{aligned}$$

We then multiply the equations by u and w respectively and integrate over $\mathbb{R}$ to get

$$\int_{\mathbb{R}} u_t(t,x)u(t,x) + cu_x(t,x)u(t,x)dx = \int_{\mathbb{R}} w(t,x)u(t,x)dx$$
$$\int_{\mathbb{R}} w_t(t,x)w(t,x) - cw_x(t,x)w(t,x) = \int_{\mathbb{R}} f(t,x)w(t,x)dx.$$

We now deal with the first equation. We first have by chain rule and Leibniz rule

$$\int_{\mathbb{R}} u_t(t,x)u(t,x)dx = \frac{d}{dt}\frac{1}{2}\int_{\mathbb{R}} u(t,x)^2dx.$$

The second term actually integrates to zero due to the fast decay at ∞ assumption:

$$\begin{aligned}\int_{\mathbb{R}} u_x(t,x)u(t,x)dx &= \lim_{L\to\infty}\int_{-L}^{L} u_x(t,x)u(t,x)dx = \lim_{L\to\infty}\int_{-L}^{L}\partial_x\frac{1}{2}u(t,x)^2dx \\ &= \lim_{L\to\infty}\left(u(t,L)^2 - u(t,-L)^2\right) = 0.\end{aligned}$$

Finally the RHS can be handled with Young's inequality:

$$\int_{\mathbb{R}} w(t,x)u(t,x)dx \leq \frac{1}{2}\int_{\mathbb{R}} w(t,x)^2 + u(t,x)^2dx.$$

This results in the estimate:

$$\frac{d}{dt}\frac{1}{2}\int_{\mathbb{R}} u(t,x)^2dx \leq \frac{1}{2}\int_{\mathbb{R}} w(t,x)^2 + u(t,x)^2dx.$$

We can apply the same techniques to the equation for w to get

$$\frac{d}{dt}\frac{1}{2}\int_{\mathbb{R}} u(t,x)^2dx \leq \frac{1}{2}\int_{\mathbb{R}} w(t,x)^2 + f(t,x)^2dx.$$

Adding these two estimates together yields the inequality

$$\frac{d}{dt}\frac{1}{2}\int_{\mathbb{R}} u(t,x)^2 + w(t,x)^2dx \leq \int_{\mathbb{R}} w(t,x)^2 + u(t,x)^2dx + \frac{1}{2}\int_{\mathbb{R}} f(t,x)^2dx.$$

Notice that the LHS on the energy estimate we want to prove is $\int_{\mathbb{R}} u(t,x)^2 + w(t,x)^2dx$. Let $\phi(t) = \int_{\mathbb{R}} u(t,x)^2 + w(t,x)^2dx$. We currently have the inequality

$$\phi'(t) \leq 2\phi(t) + \frac{1}{2}\int_{\mathbb{R}} f(t,x)^2dx.$$

The desired energy estimate

$$\phi(t) \leq e^{2t}\left(\phi(0) + \int_0^t\int_{\mathbb{R}} f(s,x)^2dxds\right)$$

is a consequence of Gronwall's inequality, which we prove below. □

Lemma 5.1 (Gronwall inequality)**.** Suppose $\phi \in C^1$ satisfies

$$\phi'(t) \leq \beta(t) + \alpha\phi(t),$$

where $\beta(t) \geq 0$ for all t and $\alpha \geq 0$. Then,

$$\phi(t) \leq e^{\alpha t}\left(\int_0^t \beta(s)ds + \phi(0)\right).$$

Proof. If ϕ satisfies the above inequality, then ϕ solves the ODE:

$$\phi'(t) = \gamma(t) + \alpha\phi(t),$$

with $\gamma(t) \leq \beta(t)$. The solution to the above ODE is

$$\phi(t) = e^{\alpha t}\phi(0) + \int_0^t e^{\alpha(t-s)}\gamma(s)ds.$$

We can then bound the second term using the fact that $e^{\alpha(t-s)} \leq e^{\alpha t}$ for all $0 \leq s \leq t$ and $\gamma(s) \leq \beta(s)$ for all s:

$$\phi(t) = e^{\alpha t}\phi(0) + \int_0^t e^{\alpha(t-s)}\gamma(s)ds \leq e^{\alpha t}\left(\phi(0) + \int_0^t \beta(s)ds\right),$$

which completes the proof. □

An important consequence of the energy estimate is that solutions that decay quickly at ∞ to the wave equation are unique.

Proposition 5.2 (uniqueness of solution to wave equation)**.** A C^2 solution to (3) with fast decay at ∞ is unique.

Proof. Let $u, v \in C^2$ be two such solutions to the wave equation (3). Notice that their difference $e = u - v$ solves

$$\begin{cases} e_{tt}(t,x) - c^2 e_{xx}(t,x) = 0, & t > 0, x \in \mathbb{R} \\ e(0,x) = 0, & x \in \mathbb{R} \\ e_t(0,x) = 0, & x \in \mathbb{R} \end{cases}.$$

We apply the energy estimate to get

$$\int_{\mathbb{R}} e(t,x)^2 + (e_t(t,x) + ce_x(t,x))^2 dx \leq 0.$$

Importantly $\int_{\mathbb{R}} e(t,x)^2 dx = 0$, so $e(t,x) = 0$ for all t, x. □

6. Solving the wave equation on an interval: separation of variables

We now depart from solving the wave equation on $\mathbb{R}$ and look at the interval $(0,1)$. We are interested in solving

$$\begin{cases} u_{tt}(t,x) - c^2 u_{xx}(t,x) = 0, & t > 0, x \in (0,1) \\ u(t,0) = u(t,1) = 0, & t > 0 \\ u(0,x) = u_0(x), & x \in (0,1) \\ u_t(0,x) = g_0(x), & x \in (0,1) \end{cases} . \tag{4}$$

The boundary conditions in this case are homogenous Dirichlet boundary conditions. We now discuss separation of variables to solve (4). It follows very similarly to the heat equation.

For Dirichlet boundary conditions, like the heat equation, we seek solutions of the form:

$$u(t,x) = \sum_{k=1}^{\infty} T_k(t) \sin(k\pi x),$$

where T_k is an unknown function of time. We also compute Fourier sine series for u_0 and g_0.

$$u_0(x) = \sum_{k=1}^{\infty} a_k \sin(k\pi x)$$
$$g_0(x) = \sum_{k=1}^{\infty} b_k \sin(k\pi x).$$

We now plug u into the PDE to see that

$$\sum_{k=1}^{\infty} (T_k''(t) + (c^2 k^2 \pi^2) T_k(t)) \sin(k\pi x) = 0$$
$$\sum_{k=1}^{\infty} T_k(0) \sin(k\pi x) = \sum_{k=1}^{\infty} a_k \sin(k\pi x)$$
$$\sum_{k=1}^{\infty} T_k'(0) \sin(k\pi x) = \sum_{k=1}^{\infty} b_k \sin(k\pi x).$$

Just like the heat equation, we can now solve for each T_k by solving the following initial value problem:

$$\begin{aligned} T_k''(t) + (c^2k^2\pi^2)T_k(t) &= 0 \\ T_k(0) &= a_k \\ T_k'(0) &= b_k. \end{aligned}$$

Using techniques from ODEs, we can find the solution to this IVP as

$$T_k(t) = a_k \cos(ck\pi t) + \frac{b_k}{ck\pi}\sin(ck\pi t),$$

and the solution to (4) is

$$u(t,x) = \sum_{k=1}^{\infty}\left(a_k\cos(ck\pi t) + \frac{b_k}{ck\pi}\sin(ck\pi t)\right)\sin(k\pi x)$$

We now go through some examples that are modifications of (4).

Example 6.1 (string with springs)**.** Suppose the string we pluck has springs attached to it. The resulting PDE would be

$$(5) \qquad \begin{cases} u_{tt}(t,x) - c^2u_{xx}(t,x) + Ku(t,x) = 0, & t>0, x\in(0,1) \\ u(t,0) = u(t,1) = 0, & t>0 \\ u(0,x) = u_0(x), & x\in(0,1) \\ u_t(0,x) = g_0(x), & x\in(0,1) \end{cases},$$

where $K \geq 0$ is a spring constant. We again suppose u takes the form

$$u(t,x) = \sum_{k=1}^{\infty} T_k(t)\sin(k\pi x).$$

Plugging u into (5) leads to the following IVP for teach T_k:

$$\begin{aligned} T_k''(t) + (c^2k^2\pi^2 + K)T_k(t) &= 0 \\ T_k(0) &= a_k \\ T_k'(0) &= b_k. \end{aligned}$$

Using techniques from ODEs, we can find the solution to this IVP as

$$T_k(t) = a_k\cos(\sqrt{c^2k^2\pi^2+K}t) + \frac{b_k}{\sqrt{c^2k^2\pi^2+K}}\sin(\sqrt{c^2k^2\pi^2+K}t),$$

and the solution to (5) is

$$u(t,x) = \sum_{k=1}^{\infty}\left(a_k\cos(\sqrt{c^2k^2\pi^2+K}\ t) + \frac{b_k}{\sqrt{c^2k^2\pi^2+K}}\sin(\sqrt{c^2k^2\pi^2+K}\ t)\right)\sin(k\pi x).$$

Example 6.2 (string with springs and friction)**.** We slightly modify the last example and now add friction to the system.

$$(6)\qquad \begin{cases} u_{tt}(t,x)+\mu u_t(t,x)-c^2u_{xx}(t,x)+Ku(t,x)=0, & t>0, x\in(0,1)\\ u(t,0)=u(t,1)=0, & t>0\\ u(0,x)=u_0(x), & x\in(0,1)\\ u_t(0,x)=g_0(x), & x\in(0,1) \end{cases},$$

where $\mu \geq 0$ is a friction coefficient.

Plugging u into (6) leads to the following IVP for teach T_k:

$$\begin{aligned} T_k''(t)+\mu T_k'(t)+(c^2k^2\pi^2+K)T_k(t)&=0\\ T_k(0)&=a_k\\ T_k'(0)&=b_k. \end{aligned}$$

Using techniques from ODEs, we can find the solution to this IVP (assuming we are underdampled: $c^2k^2\pi^2+K-\mu/2\geq 0$):

$$T_k(t)=e^{-\mu t/2}c_k\cos(\sqrt{c^2k^2\pi^2+K-\mu/2}\,t)+\tilde{c}_k\sin(\sqrt{c^2k^2\pi^2+K-\mu/2}\,t),$$

where $c_k, \tilde{c}_k$ are new constants that depend on a_k, b_k, c, k, and μ.

6.1. **Other boundary conditions.** What if we change the boundary conditions to Neumann boundary conditions? Then the wave equation reads

$$(7)\qquad \begin{cases} u_{tt}(t,x)-c^2u_{xx}(t,x)=0, & t>0, x\in(0,1)\\ u_x(t,0)=u_x(t,1)=0, & t>0\\ u(0,x)=u_0(x), & x\in(0,1)\\ u_t(0,x)=g_0(x), & x\in(0,1) \end{cases}.$$

changing from dirichlet to neumann should immediately tell us to switch from sin to cos

The only thing that changes is that our eigenvalue problem for the Fourier series changes to

$$-X_k(x)''=\lambda_k X_k(x) \text{ on } (0,1),\quad X_k'(0)=X_k'(1)=0.$$

We know the solution to this problem are cosines $X_k(x)=\cos(k\pi x)$ with $\lambda_k=k^2\pi^2$ for $k=0,\ldots$, so the solution to (7) is

$$u(t,x)=\sum_{k=0}^{\infty}T_k(t)\cos(k\pi x)$$

where T_k solves

$$T_k''(t)+(c^2k^2\pi^2)T_k(t)=0,\quad T_k(0)=\tilde{a}_k,\quad T_k'(0)=\tilde{b}_k.$$

Here, we have written

$$u_0(x) = \sum_{k=0}^{\infty} \tilde{a}_k \cos(k\pi x), \quad g_0(x) = \sum_{k=0}^{\infty} \tilde{b}_k \cos(k\pi x).$$

6.2. **Change in domain.** A last fun example is what happens if we change the domain from $(0,1)$ to $(0,L)$. With Dirichlet BC, we can think of this as holding down the fret of a guitar and then plucking the string. The wave equation here is

$$\begin{cases} u_{tt}(t,x) - c^2 u_{xx}(t,x) = 0, & t > 0, x \in (0,L) \\ u(t,0) = u(t,L) = 0, & t > 0 \\ u(0,x) = u_0(x), & x \in (0,L) \\ u_t(0,x) = g_0(x), & x \in (0,L) \end{cases} . \tag{8}$$

Again the only part of the procedure that changes is the eigenvalue problem, which is now

$$-X_k(x)'' = \lambda_k X_k(x) \text{ on } (0,L), \quad X_k'(0) = X_k'(L) = 0.$$

The solutions to this eigenvalue problem are $X_k(x) = \sin\left(\frac{k\pi}{L}x\right)$ with $\lambda_k = \left(\frac{k\pi}{L}\right)^2$. The solution to (8) is

$$u(t,x) = \sum_{k=1}^{\infty} \left(a_k \cos\left(\frac{ck\pi}{L}t\right) + \frac{b_k}{ck\pi} \sin\left(\frac{ck\pi}{L}t\right) \right) \sin(k\pi x)$$

where

$$u_0(x) = \sum_{k=0}^{\infty} a_k \sin\left(\frac{k\pi}{L}x\right), \quad g_0(x) = \sum_{k=0}^{\infty} b_k \sin\left(\frac{k\pi}{L}x\right).$$

Remark 6.1 (harmonics of a string)**.** If we were to pluck a string on a guitar of length L, the various frequencies you would here would be $\{\frac{k\pi}{L}\}_{k=1}^{\infty} = \{\sqrt{\lambda_k}\}_{k=1}^{\infty}$. If we press down halfway down a guitar string, then we would here harmonics one octave higher. In general, there is a map φ that maps $(0,L)$ to a sequence of harmonics $\{\sqrt{\lambda_k}\}_{k=1}^{\infty}$. You can see that the map is one to one. That is, if the harmonics we would here are the same, then the interval is the same.

In general, this question was asked for 2D domains and the eigenvalues of the Laplacian. If we solve

$$-u_{xx} - u_{yy} = \lambda u \text{ in } \Omega, \quad u|_{\partial\Omega} = 0,$$

and map Ω to harmonics (essentially λ), is this map 1-1? This question was asked by Mark Kac in "Can One Hear the Shape of a Drum?" in 1966. The answer is no and was answered in 1992.

7. Energy estimates and uniqueness of solutions on an interval

We return to the wave equation on an interval

$$\begin{cases} u_{tt}(t,x) - c^2 u_{xx}(t,x) = f(t,x), & t>0, x\in(0,1) \\ u(t,0) = u(t,1) = 0, & t>0 \\ u(0,x) = u_0(x), & x\in(0,1) \\ u_t(0,x) = g_0(x), & x\in(0,1) \end{cases}, \tag{9}$$

and now discuss energy estimates and uniqueness. We first begin with the energy

$$E(t) = \frac{1}{2}\int_0^1 u_t(t,x)^2 + c^2 u_x(t,x)^2 dx. \tag{10}$$

Notice that this an energy in the familiar physics sense. Here u_t^2 is like mv^2 or kinetic energy and $c^2u_x(t,x)^2$ is like an energy similar to the potential energy of a stretch spring. We know from physics that energy should be conserved unless there is outside work being put into the system. In fact, the wave equation has a conservation of energy. We compute the derivative of the energy and apply Leibniz rule and chain rule:

$$\frac{d}{dt}E(t) = \int_0^1 u_t(t,x)u_{tt}(t,x) + c^2 u_x(t,x)u_{xt}(t,x)dx$$

Notice that we have u_{tt}. In order to use the wave equation, we'd like to have a $-c^2u_{xx}$ in the integrand. This necessitates integration by parts. If u is smooth, we may swap derivatives $u_{xt}(t,x) = u_{tx}(t,x)$ and then integrate the second term by parts to get

$$\int_0^1 c^2 u_x(t,x)u_{xt}(t,x)dx = -\int_0^1 c^2 u_{xx}(t,x)u_t(t,x)dx + \left(c^2u_x(t,1)u_t(t,1) - c^2u_x(t,0)u_t(t,0)\right)$$

Notice that the boundary terms go away because of the boundary condition $u(t,0) = u(t,1) = 0$. Hence,

$$\frac{d}{dt}E(t) = \int_0^1 u_t(t,x)\left(u_{tt}(t,x) - c^2u_{xx}(t,x)\right)dx = \int_0^1 u_t(t,x)f(t,x)dx.$$

If no outside work is being performed, i.e. $f=0$, then $\frac{d}{dt}E(t) = 0$ and energy is conserved. Otherwise, we can estimate the energy using standard techniques by first applying Young's inequality

$$\begin{aligned} \frac{d}{dt}E(t) = \int_0^1 u_t(t,x)f(t,x)dx. &\le \frac{1}{2}\int_0^1 u_t(t,x)^2dx + \frac{1}{2}\int_0^1 f(t,x)^2dx \\ &\le E(t) + \frac{1}{2}\int_0^1 f(t,x)^2dx \end{aligned}$$

We can then apply Gronwall's inequality to estimate

$$E(t) \leq e^t \left(E(0) + \frac{1}{2} \int_0^t \int_0^1 f(s,x)^2 dx ds \right),$$

and we can summarize our work in the following proposition.

Proposition 7.1 (energy estimate)**.** Let $u \in C^2$ be a solution to (9). Then the energy in (10) satisfies the estimate

$$E(t) \leq e^t \left(E(0) + \frac{1}{2} \int_0^t \int_0^1 f(s,x)^2 dx ds \right).$$

Moreover, if $f = 0$, we can improve the result to

$$E(t) = E(0).$$

Remark 7.1 (test function)**.** Note that after applying Leibniz rule and integration by parts, we had

$$\frac{d}{dt} E(t) = \int_0^1 u_t(t,x) \left(u_{tt}(t,x) - c^2 u_{xx}(t,x) \right) dx.$$

This tells us that u_t is the function we want to multiply the equation by in order to prove energy estimates for more complicated situations.

An important consequence of the energy estimates and stability is uniqueness, which we have seen many times in this course.

Proposition 7.2 (uniqueness of solution to wave equation on interval)**.** Let $u, v \in C^2$ be solutions to (9). Then $u = v$.

Proof. We set $w = u - v$ and see that w solves

$$\begin{cases} w_{tt}(t,x) - c^2 w_{xx}(t,x) = 0, & t > 0, x \in (0,1) \\ w(t,0) = w(t,1) = 0, & t > 0 \\ u(0,x) = 0, & x \in (0,1) \\ w_t(0,x) = 0, & x \in (0,1) \end{cases}.$$

We then apply the energy estimate to see that $E(t) = E(0) = 0$ for all t. Hence, $w_x(t,x) = 0$ for all t, x. Since we have $w(t,0) = 0$, we can conclude that $w(t,x) = 0$ for all x, t. □

Remark 7.2 (lack of maximum principle)**.** The other concept of stability we have seen in this course in addition to energy estimates is the maximum principle. We note

that the wave equation does not have a maximum principle. To see this, consider $u_0(x) = \sin(\pi x), g_0(x) = 0, f(t,x) = 0$. Then the solution to (9) is

$$u(t,x) = \cos(c\pi t)\sin(\pi x),$$

which takes on positive and negative values even though the initial condition is nonnegative. Hence, the wave equation does not have a maximum principle.

8. Numerics for the wave equation

We now begin the discussion of numerics for the wave equation on an interval in (9)

$$\begin{cases} u_{tt}(t,x) - c^2 u_{xx}(t,x) = f(t,x), & t > 0, x \in (0,1) \\ u(t,0) = u(t,1) = 0, & t > 0 \\ u(0,x) = u_0(x), & x \in (0,1) \\ u_t(0,x) = g_0(x), & x \in (0,1) \end{cases}.$$

The first method will consider a second order finite difference for the second derivative:

$$u_{tt}(t_n, x_j) = \frac{u(t_{n+1}, x_j) - 2u(t_n, x_j) + u(t_{n-1}, x_j)}{\tau^2} + \mathcal{O}(\tau^2).$$

Notice that this approximation is at the point (t_n, x_j), so we want to match the approximations for f and u_{xx} also at (t_n, x_j). The resulting iteration for the discrete solution $\mathbf{U}_j^n$ is

$$\frac{\mathbf{U}_j^{n+1} - 2\mathbf{U}_j^n + \mathbf{U}_j^{n-1}}{\tau^2} - c^2 \frac{\mathbf{U}_{j+1}^n - 2\mathbf{U}_j^n + \mathbf{U}_{j-1}^n}{h^2} = \mathbf{f}_j^n. \tag{11}$$

At the boundaries, we use the typical technique and just set

$$\mathbf{U}_0^{n+1} = \mathbf{U}_N^{n+1} = 0.$$

This iteration is has second order truncation error, which can be shown using Taylor expansions at (t_n, x_j).

Proposition 8.1 (consistency and truncation error estimate)**.** Let $\mathbf{u}_j^n = u(t_n, x_j)$ where u solves (9) exactly. Plugging $\mathbf{u}$ into the the above iteration in (11) leads to

$$\frac{\mathbf{u}_j^{n+1} - 2\mathbf{u}_j^n + \mathbf{u}_j^{n-1}}{\tau^2} - c^2 \frac{\mathbf{u}_{j+1}^n - 2\mathbf{u}_j^n + \mathbf{u}_{j-1}^n}{h^2} = \mathbf{f}_j^n + \boldsymbol{\tau}_j^n,$$

where the truncation error satisfies

$$|\boldsymbol{\tau}_j^n| \leq C(\tau^2 |u_{tt}|_{max} + h^2 |u_{xx}|_{max}).$$

Proof. Taylor expansions at (t_n, x_j). □

8.1. **Starting the iteration.** Notice that if we substitute $n = 0$ into the iteration in (11), we get

$$\frac{\mathbf{U}_j^1 - 2\mathbf{U}_j^0 + \mathbf{U}_j^{-1}}{\tau^2} - c^2 \frac{\mathbf{U}_{j+1}^0 - 2\mathbf{U}_j^0 + \mathbf{U}_{j-1}^0}{h^2} = \mathbf{f}_j^0.$$

Notice that we need access to $\mathbf{U}^{-1}$. This is not part of the computational grid. We need to resort to a different method to kick start the iteration that is second order. The idea is to use a Taylor expansion of the exact solution:

$$u(\tau, x_j) = u(0, x_j) + \tau u_t(0, x_j) + \frac{\tau^2}{2} u_{tt}(0, x_j)$$

We can then substitute the initial conditions to get

$$u(\tau, x_j) = u_0(x_j) + \tau g_0(x_j) + \frac{\tau^2}{2} u_{tt}(0, x_j).$$

For the last term, we use the PDE to write $u_{tt}(0, x_j) = f(0, x_j) + c^2 u_{xx}(0, x_j) = f(0, x_j) + c^2 \partial_x^2 u_0(x_j)$, and

$$u(\tau, x_j) = u_0(x_j) + \tau g_0(x_j) + \frac{\tau^2}{2} \left(f(0, x_j) + c^2 \partial_x^2 u_0(x_j) \right).$$

At the discrete level, the above equation becomes

$$\mathbf{U}_j^1 = \mathbf{U}_j^0 + \tau \mathbf{g}_j + \frac{\tau^2}{2} \left(\mathbf{f}_j^0 + c^2 \left(\frac{\mathbf{U}_{j+1}^n - 2\mathbf{U}_j^n + \mathbf{U}_{j-1}^n}{h^2} \right) \right).$$

8.2. **Vector notation.** We can now write the whole scheme using vector notation. Recall the matrix $\mathbf{A}^h$ that is defined by computing the negative second finite difference

$$(\mathbf{A}^h \mathbf{v})_j = -\frac{\mathbf{v}_{j+1} - 2\mathbf{v}_j + \mathbf{v}_{j-1}}{h^2}.$$

We used this matrix in the heat equation. The finite difference scheme we derived has two components.

- **Initialization**: We set

$$\mathbf{U}_j^0 = u_0(x_j), \quad \mathbf{U}^1 = \mathbf{U}^0 + \tau \mathbf{g} + \frac{\tau^2}{2} \left(\mathbf{f}^0 - c^2 \mathbf{A}^h \mathbf{U}^0 \right)$$

- **Iteration**: To compute $\mathbf{U}^{n+1}$ for $n > 0$, we set

$$\frac{\mathbf{U}^{n+1} - 2\mathbf{U}^n + \mathbf{U}^{n-1}}{\tau^2} + c^2 \mathbf{A}^h \mathbf{U}^n = \mathbf{f}^n.$$

expect second finite difference in here

8.3. **Stability of the method.** We know the scheme in (11) is consistent. The other ingredient to prove convergence is stability. We first the question of stability in terms of a von Neumann analysis.

To do a von Neumann analysis for a scheme with multiple steps, we think of the application of one step of the scheme as the application of an operator B. That is $\mathbf{U}^n = B\mathbf{U}^{n-1}$ and $\mathbf{U}^{n+1} = B\mathbf{U}^n = B^2\mathbf{U}^{n-1}$. To compute the symbol of B, we replace $\mathbf{U}_j^{n-1} = \mathbf{v}_j = e^{i\xi hj}$, $\mathbf{U}_j^n = S(\xi h)\mathbf{v}_j$, and $\mathbf{U}_j^{n+1} = S(\xi h)^2\mathbf{v}_j$. We then use the iteration in (11):

$$\frac{\mathbf{U}_j^{n+1} - 2\mathbf{U}_j^n + \mathbf{U}_j^{n-1}}{\tau^2} - c^2\frac{\mathbf{U}_{j+1}^n - 2\mathbf{U}_j^n + \mathbf{U}_{j-1}^n}{h^2}$$

and do the substitutions

$$\frac{S(\xi h)^2\mathbf{v}_j - 2S(\xi h)\mathbf{v}_j + \mathbf{v}_j}{\tau^2} - c^2 S(\xi h)\frac{\mathbf{v}_{j+1} - 2\mathbf{v}_j + \mathbf{v}_{j-1}}{h^2} = 0$$

We now multiply by τ^2

$$S(\xi h)^2\mathbf{v}_j - 2S(\xi h)\mathbf{v}_j + \mathbf{v}_j - \frac{c^2\tau^2}{h^2}S(\xi h)\left(\mathbf{v}_{j+1} - 2\mathbf{v}_j + \mathbf{v}_{j-1}\right) = 0$$

Recall that

$$\mathbf{v}_{j+1} - 2\mathbf{v}_j + \mathbf{v}_{j-1} = \mathbf{v}_j\left(e^{i\xi h} + e^{-i\xi h} - 2\right) = \mathbf{v}_j 2\left(\cos(\xi h) - 1\right)$$

Grouping like terms and dividing out the $\mathbf{v}_j$ we are left with

$$S(\xi h)^2 + 2(\lambda^2(1 - \cos(\xi h)) - 1)S(\xi h) + 1 = 0,$$

where $\lambda = \frac{c\tau}{h}$. Since $S(\xi h)$ solves the above quadratic equation, we use the quadratic formula to write

$$S(\xi h) = \frac{-b}{2} \pm \frac{\sqrt{b^2 - 4}}{2}.$$

In order to show the symbol satisfies $|S(\xi h)| \leq 1$, we first need to determine whether the symbol has an imaginary part. From now on, we assume we satisfy the CFL condition

$$\lambda \leq 1 \text{ or equivalently } \tau \leq \frac{h}{c}.$$

We now build up inequalities on b.

$$\begin{aligned} -1 \leq & -\cos(\xi h) \leq 1 \\ 0 \leq & 1-\cos(\xi h) \leq 2 \\ 0 \leq & \lambda^2(1-\cos(\xi h)) \leq 2\lambda^2 \\ -1 \leq & \lambda^2(1-\cos(\xi h))-1 \leq 2\lambda^2-1 \\ -2 \leq & 2\left[\lambda^2(1-\cos(\xi h))-1\right] \leq 4\lambda^2-2 \end{aligned}$$

We then use the fact that $\lambda \leq 1$ to conclude $-2 \leq b \leq 2$. Hence, $b^2-4 \leq 0$, and

$$S(\xi h) = \frac{-b}{2} + i\frac{\sqrt{4-b^2}}{2}.$$

We now compute the modulus squared of the symbol to get

$$|S(\xi h)|^2 = \frac{b^2}{4} + \frac{4-b^2}{4} = 1.$$

We summarize our work in the following proposition,

Proposition 8.2 (stability of explicit method)**.** Suppose $\tau \leq \frac{h}{c}$. The iteration in (11) is stable in the sense that the symbol satisfies $|S(\xi h)| \leq 1$.

8.4. **von Neumann analysis on interval.** Recall that the von Neumann analysis looked at an infinite grid. In fact, we can repeat the same analysis on a grid of an interval. All we need are the eigenvectors of $\mathbf{A}^h$. Recall we had the following lemma in order to prove a discrete Poincare inequality.

Lemma 8.1 (eigenvalues of $\mathbf{A}^h$)**.** The eigenvalues of $\mathbf{A}^h \in \mathbb{R}^{(N-1)\times(N-1)}$ are

$$\lambda_k = \frac{4}{h^2}\sin^2\left(\frac{k\pi h}{2}\right), \quad k = 1, \ldots, N-1$$

with eigenvectors

$$\mathbf{v}_i^k = \sin(k\pi x_i).$$

eigenvalues of A^h

Proof. We just need to verify

$$\left(-\mathbf{v}_{i-1}^k + 2\mathbf{v}_i^k - \mathbf{v}_{i+1}^k\right) = 4\sin^2\left(\frac{k\pi h}{2}\right)\mathbf{v}_i^k.$$

We first split the LHS into two parts

$$-\mathbf{v}_{i-1}^k + 2\mathbf{v}_i^k - \mathbf{v}_{i+1}^k = (\mathbf{v}_i^k - \mathbf{v}_{i-1}^k) + (\mathbf{v}_i^k - \mathbf{v}_{i+1}^k).$$

The first term is

$$\mathbf{v}_i^k - \mathbf{v}_{i-1}^k = \sin(k\pi x_i) - \sin(k\pi x_{i-1}) = \sin(k\pi x_i) - \sin(k\pi(x_i - h)).$$

We then use a angle summation formula for sine:

$$\sin(k\pi(x_i - h)) = \sin(k\pi x_i)\cos(k\pi h) - \cos(k\pi x_i)\sin(k\pi h),$$

so

$$\mathbf{v}_i^k - \mathbf{v}_{i-1}^k = \sin(k\pi x_i)(1 - \cos(k\pi h)) + \cos(k\pi x_i)\sin(k\pi h).$$

We can apply the same techniques to the second term

$$\mathbf{v}_i^k - \mathbf{v}_{i+1}^k = \sin(k\pi x_i)(1 - \cos(k\pi h)) - \cos(k\pi x_i)\sin(k\pi h).$$

Adding both terms together leads to

$$-\mathbf{v}_{i-1}^k + 2\mathbf{v}_i^k - \mathbf{v}_{i+1}^k = 2(1 - \cos(k\pi h))\sin(k\pi x_i) = 2(1 - \cos(k\pi h))\mathbf{v}_i^k.$$

A useful double angle formula for cosine is

$$2(1 - \cos(k\pi h)) = 4\sin^2\left(\frac{k\pi h}{2}\right)$$

□

Once we have the eigenvectors and eigenvalues, we can repeat the von Neumann analysis but use the eigenvalues of $\mathbf{A}^h$. We write the wave equation iteration in matrix notation:

$$\frac{\mathbf{U}^{n+1} - 2\mathbf{U}^n + \mathbf{U}^{n-1}}{\tau^2} + c^2\mathbf{A}^h\mathbf{U}^n = 0$$

and rearrange

$$\mathbf{U}^{n+1} = (2\mathbf{I} - c^2\tau^2\mathbf{A}^h)\mathbf{U}^n - \mathbf{U}^{n-1}$$

We now express

$$\mathbf{U}^{n-1} = \sum_{k=1}^{N-1} a^k\mathbf{v}^k$$

We now seek solutions of the form $a_k^n\mathbf{v}^k$, where $\mathbf{v}^k$ is a single eigenvector of $\mathbf{A}^h$. The iteration now reads

$$a_k^{n+1}\mathbf{v}^k = (2\mathbf{I} - c^2\tau^2\mathbf{A}^h)a_k^n\mathbf{v}^k - a_k^{n-1}\mathbf{v}^k.$$

Notice that $(2\mathbf{I} - c^2\tau^2\mathbf{A}^h)\mathbf{v}^k = (2 - c^2\tau^2\lambda_k)$, so

$$a_k^{n+1}\mathbf{v}^k = (2 - c^2\tau^2\lambda_k)a_k^n\mathbf{v}^k - a_k^{n-1}\mathbf{v}^k.$$

Hence, we have an iteration for the coefficient:

$$a_k^{n+1} = (2 - c^2\tau^2\lambda_k)a_k^n - a_k^{n-1}.$$

The dynamics of this iteration are determined by the roots of the characteristic polynomial

$$s^2 - (2 - c^2\tau^2\lambda_k)s + 1 = 0$$

In order for the iteration to be stable, we want the roots to have complex modulus less than or equal to 1. We again solve for s using the quadratic formula

$$s = \frac{-b}{2} \pm \frac{\sqrt{b^2 - 4}}{2}$$

where $b = (2 - c^2\tau^2\lambda_k)$. Notice $0 \leq \lambda_k \leq \frac{4}{h^2}$, so

$$2 - \frac{4c^2\tau^2}{h^2} \leq b \leq 2.$$

If we enforce the CFL condition, $c\tau \leq h$, then $-2 \leq b \leq 2$. We then repeat the arguments of the previous section to get that $|s| \leq 1$.

Remark 8.1 (von Neumann analysis)**.** You'll notice that the analysis on the finite grid was essentially the same as the von Neumann analysis that we have studied so far. To adapt the von Neumann analysis to a finite grid in general, one needs to find the eigenvectors and eigenvalues of the relevant discrete operator matrices and then repeat the arguments above. Often the result agrees with the von Neumann analysis on an infinite grid.

8.5. **Energy conserving method for wave equation.** The last section we look at is a derivation of a method that preserves energy. We first look at the continuous problem to draw some inspiration.

Recall that a solution to the wave equation also solved the following system of transport equations

$$\begin{cases} u_t + cu_x = w \\ w_t - cw_x = f \end{cases},$$

which showed us that we'd expect an explicit method to satisfy a CFL condition $\tau \leq h/c$.

The next method will solve the wave equation by looking at the following system of equations

$$\begin{cases} u_t = v \\ v_t - c^2 u_{xx} = f \end{cases}, \tag{12}$$

which has the structure of a heat equation, which is familiar to us. For the rest of the discussion, we'll take $f = 0$.

We shall first rederive conservation of energy from the system in (12). We multiply the second equation by v and integrate over 0 to 1 to get

$$\int_0^1 v_t(t,x)v(t,x) - c^2 u_{xx}(t,x)v(t,x)dx = 0.$$

Recall that we have the standard identity

$$\int_0^1 v_t(t,x)v(t,x)dx = \frac{d}{dt}\frac{1}{2}\int_0^1 v(t,x)^2dx = \frac{d}{dt}\frac{1}{2}\int_0^1 u_t(t,x)^2dx,$$

which is the first part of the energy for the wave equation. Notice that the leftover term we have is

$$-\int_0^1 c^2 u_{xx}(t,x)v(t,x)dx,$$

which we would like to rewrite in terms of the second portion of the energy $\int_0^1 c^2|u_x|^2$. In order to move one derivative from u, we integrate by parts and use the homogenous Dirichlet boundary conditions:

$$\begin{aligned}-\int_0^1 c^2 u_{xx}(t,x)v(t,x)dx &= \int_0^1 c^2 u_x(t,x)v_x(t,x)dx - c^2u_x(t,1)v(t,1) + c^2u_x(t,0)v(t,0)\\ &= \int_0^1 c^2 u_x(t,x)v_x(t,x)dx\end{aligned}$$

Notice that the first equation in (12) tells us $v_x(t,x) = u_{tx}(t,x) = u_{xt}(t,x)$, so

$$\int_0^1 c^2 u_x(t,x)v_x(t,x)dx = \int_0^1 c^2 u_x(t,x)u_{xt}(t,x)dx = \frac{d}{dt}\frac{1}{2}\int_0^1 c^2 u_x(t,x)^2dx.$$

We then have

$$\frac{d}{dt}\left[\frac{1}{2}\int_0^1 u_t(t,x)^2dx + \frac{1}{2}\int_0^1 c^2u_x(t,x)^2dx\right] = 0,$$

which is the desired conservation of energy.

8.5.1. *Derivation of method.* The reason we recapped the derivation of conservation of energy is that arguments in the continuous problem serve as a guide for the design of a numerical method.

To achieve the goal of energy conservation we recall that Crank-Nicholson conserved the energy law exactly for the heat equation. As a result, we'll use Crank-Nicholson. To first discretize the second equation at the time point $t_{n+1/2} = \frac{t_{n+1}+t_n}{2}$:

$$\frac{\mathbf{V}^{n+1}-\mathbf{V}^n}{\tau} + c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+\mathbf{U}^n}{2}\right) = 0$$

where $\mathbf{V}^n$ is the variable that is a stand in for an approximation of u_t, we'll decide soon what $\mathbf{V}^n$ should be. Mimicking the proof of energy conservation from the

continuous problem and recalling the quadratic identity $(a-b)(a+b) = a^2 - b^2$, we take the dot product of the above equation by $h\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2}$ to get

$$h\frac{\mathbf{V}^{n+1}-\mathbf{V}^{n-1}}{\tau}\cdot\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2} + hc^2\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2}\cdot\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+\mathbf{U}^n}{2}\right) = 0.$$

We then use the quadratic identity $(a-b)(a+b) = a^2 - b^2$ to simplify the first term

$$h\frac{\mathbf{V}^{n+1}-\mathbf{V}^n}{\tau}\cdot\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2} = \frac{1}{2\tau}\left(\|\mathbf{V}^{n+1}\|_{2,h}^2 - \|\mathbf{V}^n\|_{2,h}^2\right).$$

We must now make a decision of how to discretize the first equation $u_t = v$. We have left $\mathbf{V}^n$ undecided so far. In order to follow the proof of the continuous problem, we want the discrete u_t to match what is multiplying the discrete u_{xx} (or $\mathbf{A}^h\frac{\mathbf{U}^{n+1}+\mathbf{U}^n}{2}$) in the equation for $\mathbf{V}$. Hence, we discretize $u_t = v$ with again a Crank-Nicholson type approximation.

$$\frac{\mathbf{U}^{n+1}-\mathbf{U}^n}{\tau} = \frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2}.$$

Hence, the second term now simplifies using again the quadratic identity

$$\begin{aligned} hc^2\left(\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2}\right)\cdot\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+\mathbf{U}^n}{2}\right) &= hc^2\left(\frac{\mathbf{U}^{n+1}-\mathbf{U}^n}{\tau}\right)\cdot\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+\mathbf{U}^n}{2}\right) \\ &= \frac{c^2}{2\tau}\left(\|\mathbf{U}^{n+1}\|_{\mathbf{A}^h}^2 - \|\mathbf{U}^n\|_{\mathbf{A}^h}^2\right) \end{aligned}$$

To summarize, the method we have is the system

$$\begin{cases} \frac{\mathbf{U}^{n+1}-\mathbf{U}^n}{\tau} = \frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2} \\ \frac{\mathbf{V}^{n+1}-\mathbf{V}^n}{\tau} + c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+\mathbf{U}^n}{2}\right) = 0 \end{cases},$$

which satisfies the following discrete conservation of energy relation:

$$\|\mathbf{V}^{n+1}\|_{2,h}^2 + c^2\|\mathbf{U}^{n+1}\|_{\mathbf{A}^h}^2 = \|\mathbf{V}^n\|_{2,h}^2 + c^2\|\mathbf{U}^n\|_{\mathbf{A}^h}^2.$$

Recall that we have that

$$\|\mathbf{V}^{n+1}\|_{2,h}^2 \approx \int_0^1 u_t(t_{n+1},x)^2dx, \quad c^2\|\mathbf{U}^{n+1}\|_{2,h}^2 \approx c^2\int_0^1 u_x(t_{n+1},x)^2dx,$$

so the discrete conservation of energy law above is precisely a discrete analog of the conservation of energy for the wave equation.

To implement the method, one can implement the system above, or one can simplify the method in terms of just $\mathbf{U}^n$. We begin by writing the second equation of

the scheme

$$\frac{\mathbf{V}^{n+1}-\mathbf{V}^n}{\tau}+c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+\mathbf{U}^n}{2}\right)=0$$
$$\frac{\mathbf{V}^{n}-\mathbf{V}^{n-1}}{\tau}+c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n}+\mathbf{U}^{n-1}}{2}\right)=0.$$

Our next goal is to somehow write the scheme in terms of averages $\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2}$ in order to use the first equation of the scheme and write everything in terms of $\mathbf{U}^n$. We do this by adding both equations above and dividing by 2:

$$\frac{\mathbf{V}^{n+1}-\mathbf{V}^n}{2\tau}+\frac{\mathbf{V}^{n}-\mathbf{V}^{n-1}}{2\tau}+c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+2\mathbf{U}^n+\mathbf{U}^{n-1}}{4}\right)=0$$

To write in terms of $\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2}$, we regroup the first two terms:

$$\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2\tau}-\frac{\mathbf{V}^{n}+\mathbf{V}^{n-1}}{2\tau}+c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+2\mathbf{U}^n+\mathbf{U}^{n-1}}{4}\right)=0$$

Notice that

$$\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2\tau}=\frac{\mathbf{U}^{n+1}-\mathbf{U}^n}{\tau^2}$$
$$\frac{\mathbf{V}^{n}+\mathbf{V}^{n-1}}{2\tau}=\frac{\mathbf{U}^{n}-\mathbf{U}^{n-1}}{\tau^2}.$$

Hence,

$$\underbrace{\frac{\mathbf{U}^{n+1}-\mathbf{U}^n}{\tau^2}-\frac{\mathbf{U}^{n}-\mathbf{U}^{n-1}}{\tau^2}}_{=\frac{\mathbf{U}^{n+1}-2\mathbf{U}^n+\mathbf{U}^{n-1}}{\tau^2}}+c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+2\mathbf{U}^n+\mathbf{U}^{n-1}}{4}\right)=0$$

and we have

$$\frac{\mathbf{U}^{n+1}-2\mathbf{U}^n+\mathbf{U}^{n-1}}{\tau^2}+c^2\mathbf{A}^h\left(\frac{\mathbf{U}^{n+1}+2\mathbf{U}^n+\mathbf{U}^{n-1}}{4}\right)=0.$$

Notice that the first term is a second finite difference to approximate $u_{tt}(t_n,x)$ and the second term is a time averaged approximation of $-u_{xx}(t_n,x)$, so we have a second order consistent method.

All of our work can now be summarized in the following proposition,

Proposition 8.3 (energy conserving method for the wave equation)**.** Let $\mathbf{U}^n$ be a sequence of grid functions that satisfy the iteration

$$\frac{\mathbf{U}^{n+1} - 2\mathbf{U}^n + \mathbf{U}^{n-1}}{\tau^2} + c^2 \mathbf{A}^h \left(\frac{\mathbf{U}^{n+1} + 2\mathbf{U}^n + \mathbf{U}^{n-1}}{4} \right) = 0.$$

Then, this method is second order consistent with truncation error

$$|\boldsymbol{\tau}_j^n| \leq C \left(\tau^2 |u_{tttt}|_{max} + \tau^2 |u_{xxtt}|_{max} + h^2 \tau^2 |u_{xxxx}|_{max} \right),$$

and the method conserves energy in the sense that

$$\|\mathbf{V}^{n+1}\|_{2,h}^2 + c^2 \|\mathbf{U}^{n+1}\|_{\mathbf{A}^h}^2 = \|\mathbf{V}^n\|_{2,h}^2 + c^2 \|\mathbf{U}^n\|_{\mathbf{A}^h}^2,$$

where $\frac{\mathbf{V}^{n+1}+\mathbf{V}^n}{2} = \frac{\mathbf{U}^{n+1}-\mathbf{U}^n}{\tau}$.

Proof. Energy conservation is a consequence of the arguments made above. The truncation error follows from standard Taylor arguments centered at (t_n, x_j). The only additional term you have to deal with is to show for any $f \in C^2$:

$$\frac{f(t_{n+1}) + 2f(t_n) + f(t_{n-1})}{4} = f(t_n) + \mathcal{O}(\tau^2).$$

□