Theory of Finite Elements

Ansh Desai
adesai@udel.edu

May 27, 2025

Contents

1 Day 1: Ordinary Differential Equations

1.1 Lecture 1: Introduction and Theory

1.1.1 Projectile Motion

Consider the trajectory of a projectile with launch angle α and launch speed S_0. Then, the projectile at time t has horizontal and vertical position $x(t), y(t)$ and velocity $v_x(t), v_y(t)$ such that

$$\frac{d}{dt}x(t) = v_x(t), \quad \frac{d}{dt}v_x(t) = 0$$
$$\frac{d}{dt}y(t) = v_y(t), \quad \frac{d}{dt}v_y(t) = -g$$

where g is the graviational constant. We precribe the initial conditions $x(0) = x_0$, $y(0) = y_0$, $v_x(0) = v_{x0}$, $v_y(0) = y_{y0}(0)$. This can easily be solved by hand

$$x(t) = x_0 + v_{x,0}t, \quad y(t) = y_0 + v_{y,0}t - \frac{1}{2}gt^2.$$

The typical associated optimization problem is to find α such that the horizontal distance travelled is maximized, where $v_{x,0} = S_0\cos(\alpha)$, $v_{y,0} = S_0\sin(\alpha)$. Observe that we first want to find $t^* > 0$ such that $y(t^*) = 0$. Indeed, we have that

$$0 = y_0 + S_0\sin(\alpha)t - \frac{1}{2}gt^2$$
$$\implies t^* = \frac{S_0\sin\alpha + \sqrt{S_0^2\sin^2\alpha + 2gy_0}}{g}.$$

Thus,

$$D(\alpha) = x(t^*) = x_0 + S_0\cos\alpha\frac{S_0\sin\alpha + \sqrt{S_0^2\sin^2\alpha + 2gy_0}}{g}.$$

A necessary condition for optima is $\frac{d}{d\alpha}D(\alpha) = 0$. For $y_0 = 0$,

$$0 = \frac{2S_0^2}{g}(-\sin^2\alpha + \cos^2\alpha) \implies \sin^2\alpha = \cos^2\alpha$$

and therefore $\alpha = 45$ degrees.

1.1.2 Drag

Air resistance leads to a force acting on a projectile opposing the movement

$$F = -\frac{1}{2} m\mu \|v\| v$$

where $\|v\| = \sqrt{v_x^2 + v_y^2}$ and $v = v_x + v_y$. This leads to

$$\frac{d}{dt}x(t) = v_x(t), \quad \frac{d}{dt}v_x(t) = -\frac{1}{2}\mu\|v\|v_x$$

$$\frac{d}{dt}y(t) = v_y(t), \quad \frac{d}{dt}v_y(t) = -g - \frac{1}{2}\mu\|v\|v_y$$

with initial conditions $x(0) = x_0$, $y(0) = y_0$, $v(x, 0) = v_{x,0}$, and $v(y, 0) = v_{y,0}$. This ODE does not have a closed form solution. We therefore need to approximate solutions.

1.1.3 Initial Value Problems

Definition 1.1.1. *An* ***initial value problem*** *is the task to find $x : I \to \mathbb{R}^d$ such that*

$$\frac{d}{dt}x(t) = F(t, x(t)), \quad x(t_0) = x_0$$

for given initial value $x_0 \in \mathbb{R}^d$ and source $F : I \times \mathbb{R}^d \to \mathbb{R}^d$.

Theorem 1.1.2 (Cauchy-Peano)**.** *Let F be a continuous function. Then, provided I is sufficiently small, there exists a solution to the IVP.*

Notice that this only guarantees existence and not uniqueness. For uniqness, we must have additional regularity on F.

Definition 1.1.3. *The function $F : I \times \mathbb{R}^d \to \mathbb{R}^d$ is said to be* ***uniformly Lipschitz*** *if there exists $L > 0$ such that*

$$\|F(t, x) - F(t, y)\| \leq L\|x - y\|, \quad \forall\, x, y \in \mathbb{R}^d.$$

Theorem 1.1.4 (Picard-Lindelöf)**.** *If F also satisfies a uniform Lipschitz condition, then the solution to the IVP is unique.*

It is important to note the following:

- The interval may be small.

  Example 1.1.5. $\frac{d}{dt}x(t) = (x(t))^2$ such that $x(1) = 1$ has solution $y = 1/(2 - t)$ so $I = (-\infty, 2)$. We have **finite-time blowup**.

- The results are strict.

  Example 1.1.6. $\frac{d}{dt}x(t) = 2\sqrt{x(t)}$ such that $x(0) = 0$ has infinitely many solutions. Fix $c > 0$ and set

  $$x(t) = \begin{cases} 0 & t \leq c \\ (t - c)^2 & t > c. \end{cases}$$

- Even when a unique solution exists, it can be very hard to solve the system for an explicit representation.

  Example 1.1.7. The Lorenz system is a "simple weather model". We want to find $x(t), y(t), z(t)$ solving

  $$\frac{dx}{dt} = \sigma(y - x), \quad \frac{dy}{dt} = x(\rho - z), \quad \frac{dz}{dt} = xy - \beta z$$

  with σ, β, ρ given.

1.2 Lecture 2: Time-Stepping Methods

1.2.1 Euler's Method

Assuming that a solution exists on $[t_0, T]$ to the problem

$$\frac{d}{dt}x(t) = F(t, x(t)), \quad x(t_0) = x_0,$$

let us try to numerically approximate it. We consider a discretization $t_k = t_0 + \tau k$ for $k = 0, 1, \ldots, N$, where $\tau = \frac{T-t_0}{N}$. We want to find a discrete approximation $\{y^n\}_{n=0}^N$ of $x(t)$ with $y^n \approx x(t_n)$ for $n = 0, 1, \ldots, N$. The idea is to approximate the differentiable operator by a difference operator:

$$\frac{d}{dt}x(t_n) \approx \frac{x(t_{n+1}) - x(t_n)}{t_{n+1} - t_n}$$

which implies

$$\frac{y^{n+1} - y^n}{\tau} = F(t_n, y^n).$$

Definition 1.2.1 (Euler's Method). *Construct a sequence of approximations $\{y^n\}_{n=0}^N$ as follows:*

$$y^{n+1} = y^n + \tau F(t_n, y^n)$$

where $y^0 = x_0$.

This is an explicit time-marching procedure as the RHS only depends on t_n and y^n. How good is this approximation?

Definition 1.2.2. *We define the truncation error*

$$\pi^n = \frac{x(t_{n+1}) - x(t_n)}{\tau} - F(t_n, x(t_n)).$$

By Taylor's Theorem,

$$x(t_{n+1}) = x(t_n) + \frac{d}{dt}x(t_n)(t_{n+1} - t_n) + \frac{1}{2}\frac{d^2}{dt^2}x(\xi_n)(t_{n+1} - t_n)^2$$

for some $\xi_n \in (t_n, t_{n+1})$. Using that $F(t_n, x(t_n)) = \frac{d}{dt}x(t_n)$ and substituting into the scheme,

$$\pi^n = \frac{x(t_n) + \frac{d}{dx}x(t_n)\tau + \frac{1}{2}\frac{d^2}{dx^2}x(\xi_n)\tau^2 - x(t_n)}{\tau} - \frac{d}{dt}x(t_n) = \frac{1}{2}\tau\frac{d^2}{dt^2}x(\xi_n).$$

This implies the following.

Lemma 1.2.3. *The truncation error for Euler's method is given by*

$$\max_n \|\pi^n\| \leq \frac{1}{2}\tau \max_{\xi \in I}\left\|\frac{d^2}{dt^2}x(\xi)\right\|.$$

This is a first order approximation.

1.2.2 Consistency and Stability

Definition 1.2.4. *A one-step method is* ***consistent*** *with order k if*

$$\max_n \|\tau^n\| \leq C\tau^k$$

for some $C > 0$. A one-step method is ***convergent*** *with order k if for the error $e^n = x(t_n) - y^n$,*

$$\max_n \|e^n\| \leq c\tau^k$$

for some $c > 0$.

Lemma 1.2.5 (Gronwall). *Suppose we have monotone sequences $\{w_n\}, \{b_n\}$ where b_n are increasing and a constant $a > 0$ such that*

$$w_0 \leq b_0, \quad w_{n+1} \leq a \sum_{j=0}^{n} w_j + b_{n+1}, \;\; n \geq 0.$$

Then, $w_{n+1} \leq \exp((n+1)a) b_{n+1}$ for $n \geq 0$.

Proof. Set $S_{n+1} = a \sum_{j=0}^{n} w_j + b_{n+1}$. We now show $S_{n+1} \leq \exp((n+1)a) b_{n+1}$. First, we have $S_0 \leq b_0$. Assume that it holds for n, that is, $S_n \leq \exp(na) b_n$ and by assumption $w_n \leq S_n$. Then,

$$S_{n+1} - S_n = a w_n + b_{n+1} - b_n$$

$$\implies S_{n+1} \leq (1+a) S_n + b_{n+1} - b_n \leq (1+a) e^{na} b_n + b_{n+1} - b_n \leq e^{a} e^{na} b_n + e^{(n+1)a}(b_{n+1} - b_n) \leq e^{(n+1)a} b_{n+1}.$$

By induction, the result holds for all n. □

Theorem 1.2.6 (Discrete Stability of Euler's Method). *Let F be a Lipschitz continuous function with Lipschitz constant L. Let $x(t)$ be a solution to $\frac{d}{dt}x(t) = F(t, x(t))$ and $\{y^n\}$ generated by Euler's method. Then,*

$$\max_n \|e^n\| \leq \exp(LT) T \max_n \|\pi^n\|.$$

Proof. From the definition of error,

$$\begin{aligned} e^{n+1} &= x(t_{n+1}) - y^{n+1} \\ &= x(t_n) + \tau F(t_n, x(t_n)) + \tau \pi^n - y^{n+1} \\ &= x(t_n) + \tau F(t_n, x(t_n)) + \tau \pi^n - y^n - \tau F(t_n, x(t_n)) \\ &= e^n + \tau \{F(t_n, x(t_n)) - F(t_n, y^n)\} + \tau \pi^n. \end{aligned}$$

Taking the norm and applying the Lipschitz condition

$$\|e^{n+1}\| \leq \|e^n\| + \tau L \|e^n\| + \tau \|\pi^n\| = (1 + \tau L)\|e^n\| + \tau \|\pi^n\|.$$

Recurisively, we obtain

$$\|e^{n+1}\| \leq \tau L \sum_{j=0}^{n} \|e^j\| + \tau \sum_{j=0}^{n} \|\pi^j\|.$$

From the Grownall lemma with $a = \tau L$, $b_{n+1} = \tau \sum_{j=0}^{n} \|\pi^j\|$ and $w_n = \|e^n\|$, we obtain

$$\|e^{n+1}\| \leq \exp((n+1)L\tau) \tau \sum_{j=0}^{n} \|\pi^j\|.$$

From here, the result follows. □

An important principle is that consistency and stability imply convergence.

1.2.3 Explicit Runge-Kutta Methods

It is often necessary to construct time-stepping schemes that are more than first order convergent. A huge class of such schemes fit into the framework of a Runge-Kutta method.

Definition 1.2.7. *A Runge-Kutta time stepping scheme is of the form*

$$y^{n+1} = y^n + \tau \sum_{j=1}^{R} b_j K_j$$

where

$$K_1 = F(t_n, y^n), \quad K_j = F\left(t_n + \tau c_j, y^n + \tau \sum_{i=1}^{j-1} a_{ij} K_i\right), \;\; j \geq 2.$$

$$
\begin{array}{c|cccc}
c_1 & 0 & & & \\
c_2 & a_{21} & 0 & & \\
\vdots & \vdots & \vdots & \ddots & \\
c_R & a_{R1} & a_{R2} & \cdots & 0 \\
\hline
 & b_1 & b_2 & \cdots & b_R
\end{array}
$$

Thus, we need to find the coefficients $\{a_{ij}, b_j, c_j\}$ the make this scheme converge with our desired order. They are often organized in a Butcher tableau: For consistency, we require $\sum_j b_j = 1$.

Example 1.2.8. For $R = 1$, the only possible choice is $b_1 = 1$ and we have the **explicit forward Euler scheme**

$$y^{n+1} = y^n + \tau F(t_n, y^n).$$

Example 1.2.9. For $R = 2$, we have several choices. A popular choice is **Heun's method**

$$y^{n+1} = y^n + \frac{\tau}{2} \left\{ F(t_n, y^n) + F(t_n + \tau, y^n + \tau F(t_n, y^n)) \right\}.$$

Alternatively, we have a second order Euler scheme

$$y^{n+1} = y^n + \tau \left\{ F(t_n, y^n) + F(t_n + 1/2\tau, y^n + 1/2\tau F(t_n, y^n)) \right\}.$$

Example 1.2.10. For $R = 4$, the classical **Runge-Kutta method** (RK4) is the fourth order scheme

$$y^{n+1} = y^n + \frac{\tau}{6}(K_1 + 2K_2 + 2K_3 + K_4),$$

where

$$
\begin{aligned}
K_1 &= F(t_n, y^n) \\
K_2 &= F(t_n + 1/2\tau, y^n + 1/2\tau K_1) \\
K_3 &= F(t_n + 1/2\tau, y^n + 1/2\tau K_2) \\
K_4 &= F(t_n + \tau, y^n + \tau K_3).
\end{aligned}
$$

2 Day 2: Modeling

2.1 Lecture 1: Preliminaries

2.1.1 Integration by Parts

We are all familiar with the integration by parts formula in $\mathbb{R}$:

$$\int_a^b uv' dx = (uv)\Big|_a^b - \int_a^b u'v dx.$$

To extend this we need the Divergence Theorem.

Theorem 2.1.1 (Divergence Theorem). *Let $\Omega \subset \mathbb{R}^d$ be compact with smooth boundary $\partial\Omega$ and $x \in \mathbb{R}^d$. The exterior normal to Ω is denoted $n(x)$. Then, for any $F \in \mathcal{C}^1(V)$,*

$$\int_\Omega \operatorname{div} F(x) dx = \int_{\partial\Omega} F(x) \cdot n(x) dS.$$

Using this, it is easy to obtain IBP in higher dimensions.

Theorem 2.1.2 (Integration by Parts). *Let $\Omega \subset \mathbb{R}^d$ be compact with smooth boundary $\partial\Omega$, $\phi : \Omega \to \mathbb{R}$, and $v : \Omega \to \mathbb{R}^d$. Then*

$$\int_\Omega v(x) \cdot \nabla\phi(x) dx = \int_{\partial\Omega} \phi(x) v(x) \cdot n(x) dS - \int_\Omega \phi(x) \operatorname{div} v \, dx.$$

Proof. Take $F(x) = \phi(x)v(x)$ in the Divergence Theorem. Then,

$$\operatorname{div}(\phi v) = \phi \operatorname{div}(v) + v \cdot \nabla\phi.$$

Thus,

$$\int_\Omega v \cdot \nabla\phi = \int_\Omega [\operatorname{div}(\phi v) - \phi \operatorname{div}(v)]\, dx = \int_{\partial\Omega} \phi v \cdot n \, dS - \int_\Omega \phi \operatorname{div}(v) dx.$$

□

The most important identity to remember is

$$\int_\Omega v\partial_{x_i} u \, dx = \int_{\partial\Omega} uvn_i \, dS - \int_\Omega u\partial_{x_i} v \, dx.$$

From this, many formulas follow.

Theorem 2.1.3 (Green's First Identity).

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_{\partial\Omega} v\nabla u \cdot n \, dS - \int_\Omega v\Delta u \, dx$$

where $\Delta u = \operatorname{div}(\nabla u)$ *is the Laplacian.*

How do we use the formula in practice?

Example 2.1.4 (2D Integration By Parts).

$$\begin{aligned}\int_\Omega \operatorname{div}(v)\phi &= \int_\Omega \frac{\partial v_1}{\partial x_1}\phi + \int_\Omega \frac{\partial v_2}{\partial v_2} \\ &= \int_{\partial\Omega} \phi v_1 n_1 - \int_\Omega v_1 \frac{\partial\phi}{\partial x_1} + \int_{\partial\Omega} \phi v_2 n_2 - \int_\Omega v_2 \frac{\partial\phi}{\partial x_2} \\ &= \int_{\partial\Omega} \phi v \cdot n - \int_\Omega v \cdot \nabla\phi.\end{aligned}$$

Example 2.1.5 (Non-Obvious Formula).

$$\int_\Omega \nabla \times u \, dx = -\int_{\partial\Omega} u \times n \, ds.$$

Proof. Consider that

$$\nabla \times u = \begin{bmatrix} \partial_y u_3 - \partial_z u_2 \\ \partial_z u_1 - \partial_x u_3 \\ \partial_x u_2 - \partial_y u_1 \end{bmatrix}.$$

We can do integration by parts on each row and the formula follows. □

Example 2.1.6 (Another Non-Obvious Formula).

$$\int_\Omega (\nabla \times u) \cdot v \, dx = \int_{\partial\Omega} (u \times v) \cdot n \, dS + \int_\Omega u \cdot (\nabla \times v) dx.$$

2.1.2 Total Derivative

Consider the function $f(t, x(t))$. Such a function could represent a quantity being convected/transpoted by a velocity $v(x, t)$. That is, $f(x(t), t)$ is convected by the velocity $v(x) = \frac{dx}{dt}$. Therefore, what is the rate of change of f? Using the chain rule,

$$\frac{d}{dt} f(x(t), t) = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial t}$$

or using v,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x}.$$

We refer to this as the **total derivative** or **material derivative** of f. It denotes the rate of change of a quantity that is subjected to both a position and time dependent velocity field. In the multi-dimensional case, we can similarly obtain

$$\frac{d}{dt}f(x(t),t) = \frac{\partial f}{\partial t} + \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{t} = \frac{\partial f}{\partial t} + \nabla_x f \cdot v.$$

More generally, this quantity is often called the **Lie derivative of 0-forms**.

2.1.3 Integrals in Time-Dependent Domains

Recall that for continuous $f : [a,b] \to \mathbb{R}$,

$$\int_a^b f(x)dx = F(b) - F(a),$$

that is, f admits a primative F. If a, b are time-dependent, i.e., we now are on the interval $[a(t), b(t)]$, then

$$\int_{a(t)}^{b(t)} f(s)ds = F(b(t)) - F(a(t)).$$

Using the chain rule,

$$\frac{d}{dt}\left[\int_{a(t)}^{b(t)} f(s)ds\right] = f(b(t))b'(t) - f(a(t))a'(t).$$

This is often called the **Leibniz Rule** where the quantities $a'(t), b'(t)$ are **boundary velocities**. We can generalize this to the multi-dimensional setting, though we do not prove it.

Theorem 2.1.7 (Reynold's Transport Theorem). *Let $\Omega(t)$ be the domain of integration. Let $f = f(x,t)$ be scalar, vector, or tensor-valued. Then,*

$$\frac{d}{dt}\int_{\Omega(t)} f\, dV = \int_{\Omega(t)} \frac{\partial f}{\partial t}dV + \int_{\partial\Omega(t)} (v_b \cdot n) f\, dA$$

where $n(x,t)$ is the outward pointing normal and v_b is the velocity of the area element.

We could also alternative use the Divergence Theorem to reformulate Reynold's Transport Theorem as

$$\frac{d}{dt}\int_{\Omega(t)} f dx = \int_{\Omega(t)} (\frac{\partial f}{\partial t} + \operatorname{div}(vf))dx.$$

The quantity

$$\frac{\partial f}{\partial t} + \operatorname{div}(vf)$$

is referred to as the **Lie derivative of k-forms**.

2.2 Lecture 2: Conservation Laws

2.2.1 Conservation of Mass

Let $\Omega \subset \mathbb{R}^d$ with boundary $\partial\Omega$ be fixed. Let $\rho(x,t)$ denote the density (mass per unit volume) and $v(x,t)$ denote the velocity at whcih the mass moves. The total mass M in Ω is given by

$$M = \int_\Omega \rho(x,t)dx.$$

Observe that mass can only enter or leave the boundary through the boundary so

$$\frac{d}{dt}\int_{\Omega}\rho(x,t)dx = -\int_{\partial\Omega}\rho(x,t)v(x,t)\cdot n\,dS.$$

Why do we negate the left side? If $v \cdot n > 0$, then we have an outflow. If $v \cdot n < 0$, then we have an inflow. Reorganizing, we obtain

$$\int_{\Omega}\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho v)dx = 0.$$

Since this holds for arbitrary Ω, the integrand must be zero and

$$\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho v) = 0.$$

This is known as **conservation or balance of mass** as the inflow and outflow are balanced. ρ is often called a **conserved quantity** if $v \cdot n = 0$ on $\partial\Omega$. Indeed, then

$$0 = \int_{\Omega}\frac{\partial\rho}{\partial t} + \operatorname{div}(\rho v)dx = \int_{\Omega}\frac{\partial\rho}{\partial t}dx + \int_{\partial\Omega}\rho v\cdot n\,dS = \int_{\Omega}\frac{\partial\rho}{\partial t}dx$$

$$\implies \frac{\partial}{\partial t}\int_{\Omega}\rho\,dx = 0.$$

Thus,

$$\int_{\Omega}\rho(x,t)dx = \int_{\Omega}\rho(x,0)dx.$$

This means that the total mass at time t is equivalent to the total mass at the starting time $t = 0$.

2.2.2 Conservation of Momentum

For a force F acting on a point mass, Newton's law says that

$$\frac{d}{dt}(mv) = F$$

where mv is the momentum. If $\partial_t m = 0$, then we can rewrite this as

$$m\frac{dv}{dt} = F.$$

$v = v(x(t), t)$ so we interpret the acceleration $\frac{dv}{dt}$ as a total derivative. Since ρ is the mass per unit volume,

$$\rho\left(\frac{\partial v}{\partial t} + \nabla v v\right) = F.$$

But v is a vector, so we interpret

$$[\nabla v]_{ij} = \frac{\partial v_i}{\partial x_j}$$

as the Jacobian of v. More commonly, we rewrite

$$\rho\left(\frac{\partial v}{\partial t} + (v\cdot\nabla)v\right) = F.$$

The right hand side are the forces per unit volume. What types of forces act on bodies? There are two primary types.

- External forces. Act on the exterior of the body.
- Cohesive or internal forces: Generated by the interior of the body.

Example 2.2.1 (Gravity)**.** Gravity is an external force

$$F_g = \begin{bmatrix} 0 \\ 0 \\ -g \end{bmatrix}$$

where $g \approx 9.81 m/s^2$ is the graviational constant.

Internal forces are mathematically represented by tensors, that is, $F = \operatorname{div}(\sigma)$ where $\sigma \in \mathbb{R}^{d \times d}$. The divergence of a matrix is defined as

$$(\operatorname{div} \sigma)_i = \sum_{j=1}^{d} \frac{\partial}{\partial x_j} (\sigma_{ij}).$$

σ is called the **stress tensor** and characterizes the elastic/compressible nature of the substance. The precise formula of σ depends on the material or quantity of interest. It can be quite complicated, but we assume that we are given a precise formula $\sigma : \Omega \to \mathbb{R}^{n \times n}$.

Example 2.2.2 (Pressure)**.** Consider $\Omega = \Omega_1 \cup \Omega_2$ with the interface (boundary between subdomains) Γ. Let n be the normal to Γ. We can define the **traction vector** $T = \sigma n\big|_{\Gamma}$. This characterizes the force per unit area resulting from cohesive stress. Let us define the **pressure** $p = -\frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) = -\frac{1}{3}\operatorname{tr}(\sigma)$. The simplest form of the stress tensor (commonly seen in fluids) is

$$\sigma = -pI = \begin{bmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{bmatrix}.$$

In this context, $T = \sigma n = -pn$. If $p > 0$, T points opposite to n and we refer to this as **compression**. If $p < 0$, then T points in the same direction as n and we refer to this as an **attraction state**. Fluids and gasses can withstand compression but they do not support traction. In contrast, solids withstand both compression and traction.

In general therefore, we may write $F = F_e + \operatorname{div}(\sigma)$ where F_e denotes the external force. So far, we have conservation of mass and momentum:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0, \\ \rho \left[\frac{\partial v}{\partial t} + (v \cdot \nabla) v \right] - \operatorname{div}(\sigma) = F_e. \end{cases}$$

Note that conservation of momentum is not written in divergence form. To fix this, we multiply the mass law by v and add them:

$$\begin{aligned} \frac{\partial \rho}{\partial t} v + \operatorname{div}(\rho v) v &= 0 \\ \implies \frac{\partial \rho}{\partial t} v + \rho \frac{\partial v}{\partial t} + \operatorname{div}(\rho v) v + \rho (v \cdot \nabla) v - \operatorname{div}(\sigma) &= F_e \\ \implies \frac{\partial}{\partial t}(\rho v) + \operatorname{div}(\rho v v^T) - \operatorname{div}(\sigma) &= F_e. \end{aligned}$$

Here, we used that

$$\operatorname{div}(\rho v v^T) = \sum_{j=1}^{d} \partial x_j (\rho v_i v_j) = \sum_{j=1}^{d} v_i \underbrace{\partial_{x_j}(\rho v_j)}_{\operatorname{div}(\rho v)} + \rho v_j \underbrace{\partial_{x_j} v_i}_{\nabla v} = v \operatorname{div}(\rho v) + \rho \nabla v v = \operatorname{div}(\rho v) v + \rho (v \cdot \nabla) v.$$

Thus in divergence form, the conservation of mass and momentum is

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0, \\ \frac{\partial}{\partial t}(\rho v) + \operatorname{div}(\rho v v^T - \sigma) = F_e. \end{cases}$$

2.2.3 Conservation of Energy

This system is still incomplete, however. Consider the **power** of the forces applied to the system, $P = P(\rho, e)$. Here, e is the **specific internal energy** and is related to the **temperature** θ. This system does not describe the evolution of e. We assume that the gas/fluid of interest is purely theorem-mechanical. A thermo-mechanical body of fluid can only obtain energy in the form of internal and kinetic energy. In this case, the mechanical energy density

$$\varepsilon = \underbrace{\rho e}_{\text{internal}} + \underbrace{\frac{1}{2}\rho|v|^2}_{\text{kinetic}}.$$

We define

$$E = \int_\Omega \varepsilon \, dx$$

as the total thermo-mechanical energy stored in Ω. The **First Law of Thermodynamics** says that

$$\frac{dE}{dt} = \frac{d}{dt}\int_{\Omega(t)} \varepsilon \, dx = P + Q_s$$

where the rate of heat recieved by the system is

$$Q_s = \int_{\Omega(t)} \Gamma \, dx + \int_{\partial\Omega(t)} q \cdot n \, dS.$$

Here, Γ is the **source of the heat** in the bulk of Ω, q is the **flux** of the heat through $\partial\Omega$. The aformentioned power of the system is

$$P = \int_{\Omega(t)} F_e \cdot v \, dx + \int_{\partial\Omega(t)} (\sigma n) v \, dS.$$

Therefore,

$$\frac{d}{dt}\int_{\Omega(t)} \varepsilon \, dx = \int_{\Omega(t)} F_e \cdot v + \Gamma \, dx + \int_{\partial\Omega(t)} (\sigma n) \cdot v + qn \, dS.$$

Using the Reynold Transport Theorem and the Divergence Theorem,

$$\int_{\Omega(t)} \frac{\partial \varepsilon}{\partial t} + \operatorname{div}(\varepsilon v) - \operatorname{div}(\sigma v) - \operatorname{div}(q) \, dx = \int_{\Omega(t)} F_e \cdot v + \Gamma \, dx.$$

Since this holds true for any $\Omega(t)$,

$$\frac{\partial \varepsilon}{\partial t} + \operatorname{div}(\varepsilon v - \sigma v - q) = F_e \cdot v + \Gamma.$$

This is **balance of total mechanical energy**. Note that if there are no external forces, $F_e = 0$, and if there are no sources of heat, $\Gamma = 0$. So **conservation of total mechanical energy** is

$$\frac{\partial \varepsilon}{\partial t} + \operatorname{div}(\varepsilon v - \sigma v - q) = 0.$$

Now let's consider a fixed domain, i.e., $\frac{\partial \Omega}{\partial t} = 0$. If $v \cdot n|_{\partial\Omega} = 0$, $v^T \sigma n|_{\partial\Omega} = 0$, $q \cdot n|_{\partial\Omega} = 0$, then

$$\int_\Omega \frac{\partial \varepsilon}{\partial t} + \operatorname{div}(\varepsilon v - \sigma v - q) dx = 0$$

$$\implies \int_\Omega \frac{\partial \varepsilon}{\partial t} dx = -\int_\Omega (\varepsilon v \cdot n - (\sigma v) \cdot n - q \cdot n) dS = 0$$

$$\implies \int_\Omega \epsilon(t) dx = \int_[\Omega]\epsilon(0) dx$$

and so total mechanical energy is a conserved quantity.

So far, we derived an evolution equation for ε. By definition

$$\varepsilon = \rho e + \frac{1}{2}\rho\|v\|^2 \implies \rho e = \varepsilon - \frac{1}{2}\rho\|v\|^2$$

is the internal energy. Let's derive an evolution equation for ρe. We have that

$$\frac{\partial(\rho e)}{\partial t} = \frac{\partial \varepsilon}{\partial t} - \frac{\partial}{\partial t}(1/2\rho|v|^2).$$

The first term we know is

$$\frac{\partial \varepsilon}{\partial t} = -\operatorname{div}(\varepsilon v - \sigma v - q) + F_e \cdot v + \Gamma.$$

For the second term,

$$\frac{\partial}{\partial t}(1/2\rho|v|^2) = \frac{1}{2}\partial_t \rho|v|^2 + \rho v \partial_t v.$$

For the first part of this, we multiply conservation of mass by $\frac{1}{2}|v|^2$ so that

$$\partial_t \rho \frac{1}{2}|v|^2 + \operatorname{div}(\rho v)\frac{1}{2}|v|^2 = 0.$$

For the second part of this, we multiply conservation of momentum by v,

$$\rho\frac{\partial v}{\partial t} \cdot v + (\nabla v v) \cdot v - \operatorname{div}(\sigma) \cdot v = F_e \cdot v.$$

Adding both, we get an evolution of kinetic energy:

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho|v|^2\right) + \operatorname{div}\left(\frac{1}{2}|v|^2 v\right) - \operatorname{div}(\sigma v) = F_e \cdot v.$$

Now we can compute $\partial_t(\rho e)$. We have that

$$\frac{\partial}{\partial t}(\rho e) = \frac{\partial \varepsilon}{\partial t} - \frac{\partial}{\partial t}\left(\frac{1}{2}\rho|v|^2\right) = -\operatorname{div}(\rho e v) - \operatorname{div}(q) + \underbrace{\div(\sigma v) - \operatorname{div}(\sigma)}_{=\sigma:\nabla v} \cdot v + \Gamma.$$

Here, we use the double contraction

$$\sigma : \nabla v = \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_{ij}\frac{\partial v_i}{\partial x_j}.$$

Reorganizing, we express **balance of internal energy** as

$$\frac{\partial(\rho e)}{\partial t} + \operatorname{div}(\rho e v + q) = \sigma : \nabla v + \Gamma.$$

Using similar arguments we can deduce that

$$\rho\left[\frac{\partial e}{\partial t} + v \cdot \nabla e\right] + \operatorname{div}(q) = \sigma : \nabla v + \Gamma.$$

2.3 Lecture 3: Physical Models

2.3.1 Euler Equations

Assume that we have no heat sources or heat conduction so that $\Gamma = 0$, $q = 0$. Also assume that we have a diagonal stress tensor $\sigma = -pI$. Then, we obtain the system of conservation laws

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0 \\ \frac{\partial(\rho v)}{\partial t} + \operatorname{div}(\rho v v^T + Ip) = 0 \\ \frac{\partial \varepsilon}{\partial t} + \operatorname{div}[(\varepsilon + p)v] = 0. \end{cases}$$

We still have to objectify the pressure p. The simplest case is for an ideal gas.

$$p = (\gamma - 1)\rho e, \quad \rho e = \varepsilon - \frac{1}{2}\rho|v|^2,$$

with temperature $\theta = (\gamma + 1)e$ and $1 < \gamma < 5/3$ the ratio of specific heat. This is known as a **thermo-mechanical closure**. This system is of somewhat universal validity. Indeed, the formula for p is phenomonological and depends on the gas we model. A closure is a "constitutive relationship". Through this formulation, we obtain **Euler's Equations for Gas Dynamics**.

2.3.2 Acoustic Wave Equation

We could also consider the simplified system

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0 \\ \frac{\partial(\rho v)}{\partial t} + \nabla p = 0. \end{cases}.$$

Differentiating the first equation and taking the divergence of the second,

$$\begin{cases} \frac{\partial^2 \rho}{\partial t^2} + \frac{\partial}{\partial t}\operatorname{div}(\rho v) = 0 \\ \frac{\partial}{\partial t}\operatorname{div}(\rho v) = -\Delta \rho. \end{cases}$$

Combining, we have

$$\frac{\partial^2 \rho}{\partial t} - \Delta P = 0.$$

With the **isothermal closure** $p = c^2\rho$, we obtain the **acoustic wave equation**

$$\frac{\partial^2 \rho}{\partial t} - c^2 \Delta \rho = 0.$$

In the 1D case, this admits a solution as a travelling wave

$$\rho(x,t) = f(ct - x) + g(ct + x)$$

for any functions f, g.

2.3.3 Advection-Diffusion-Reaction

Finally, consider

$$\rho\frac{\partial e}{\partial t} + \rho v \cdot \nabla e + \operatorname{div}(q) = \sigma : \nabla v + \Gamma.$$

For gases, $\theta = (\gamma - 1)e$ so internal energy corresponds with temperature. For incompressible substances like solids, $e = c_p\theta$, $\rho\rho_0$ is constant (incompressible flow), and the power of viscous stress $\sigma : \nabla v$ is neglibile. Then, we get

$$\frac{\partial \theta}{\partial t} + v \cdot \nabla\theta + \frac{1}{c_p\rho_0}\operatorname{div}(q) = \frac{\Gamma}{c_p\rho_0}.$$

When $q = -k\nabla\theta$,

$$\frac{\partial \theta}{\partial t} + \underbrace{v \cdot \nabla\theta}_{\text{advection}} - \underbrace{\frac{1}{c_p\rho_0}\Delta\theta}_{\text{diffusion}} = \underbrace{\frac{\Gamma}{c_p\rho_0}}_{\text{reaction}}.$$

This is the **advection-diffusion-reaction equation**. When $v = 0, \Gamma = 0$, we obtain the **heat/diffusion equation**

$$\frac{\partial \theta}{\partial t} - \frac{k}{c_p\rho_0}\Delta\theta$$

and the quantity $k/c_p\rho_0$ is the **thermal diffusivity**.

3 Day 3: Partial Differential Equations

3.1 Lecture 1: Parabolic Equations

3.1.1 Problem Statement

The goal of this lecture is to investigate the structure of the following scalar-valued initial boundary value problem. We want to find $u : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ such that

$$\begin{aligned} \partial_t u - p\partial_{xx}u + q\partial_x u + ru = f, \quad x \in \mathbb{R}, t \in \mathbb{R}^+ \\ u(x,0) = u_0(x) \\ \lim_{x\to\pm\infty} u(x,t) = 0, \quad \forall\, t \in \mathbb{R}^+. \end{aligned}$$

Here $p, q, r \in \mathbb{R}$ are constants such that $p > 0$ and ≥ 0, and $f(\cdot, t), u_0 \in \mathbb{L}^2(\mathbb{R})$. We say that f is the **source** and u_0 is the **initial data**. The conditions $\lim_{x\to\pm\infty} u(x,t)$ are called Dirichlet boundary conditions at infinity and enforce suitable decay in our function. More generally, one could also consider a finite interval $(-L, L)$ and prescribe that $u(\pm L, t) = 0$.

Remark 3.1.1. If we replace ∂t by t, ∂_x by x, and ∂_{xx} by x^2, then we would obtain $(t - px^2 + qx + r)u = f$ (the reason for this is motivated by Fourier analysis). Since we assumed $p > 0$, the two-variate polynomial generates a **parabola** and hence it is customary to say that this problem is **parabolic**.

The first three natural questions that come to mind are

1. **Existence**: Does there exist a solution?
2. **Uniqueness**: If a solution exists, is it unique?
3. **Stability**: If a solution exists, is the solution operator $(f, u_0) \mapsto f$ continuous in some sense? In other words, if $\|(f^{(1)}, u_0^{(1)}) - (f^{(2)}, u_0^{(2)})\| < \delta$, is $\|u^{(1)} - u^{(2)}\| < \epsilon$ in some suitable metric $\|\cdot\|$?

It turns out the existence question is the most difficult one as it requires significant analysis. The easiest question is that concerning the stability/continuity. Hence, we are go- ing investigate first, the stability, then the uniqueness, and finally say a few words regarding existence.

3.1.2 Stability

The **energy method** is a principle that we will use to derive a-priori estimates for PDEs. The idea is to multiply the PDE by the solution itself and integrate by parts to obtain a useful bound. Suppose that such a solution u to our PDE exists.

Theorem 3.1.2. *Let $u(x,t)$ be a solution and $T > 0$. Then,*

$$\|u(\cdot, T)\|_{L^2(\mathbb{R})} \leq \|u_0\|_{L^2(\mathbb{R})} + \int_0^T \|f(\cdot, t)\|_{L^2(\mathbb{R})} dt.$$

Proof. Using the energy method, we multiply the PDE by u and integrate:

$$\int_{\mathbb{R}} u\partial_t u - pu\partial_{xx}u + qu\partial_x u + ru^2 = \int_{\mathbb{R}} uf.$$

We have that,

$$\begin{aligned}
\int_{\mathbb{R}} \left(u\partial_t u - pu\partial_{xx}u + qu\partial_x u + ru^2\right) &= \int_R \left(\partial_t \left(\frac{1}{2}u^2\right) - pu\partial_{xx}u + q\partial_x \left(\frac{1}{2}u^2\right) + ru^2\right) \\
&= \partial_t \int_{\mathbb{R}} \frac{1}{2}u^2 + \int_{\mathbb{R}} -pu\partial_{xx}u + \int_{\mathbb{R}} \left(q\partial_x \left(\frac{1}{2}u^2\right) + ru^2\right) \\
&= \partial_t \int_{\mathbb{R}} \frac{1}{2}u^2 - \underbrace{\Big(\lim_{N,N'\to\infty} u(x,t)\partial_x u(x,t)\Big|_{-N'}^{N}\Big)}_{0} + \int_{\mathbb{R}} p(\partial_x u)^2 + \int_{\mathbb{R}} \left(q\partial_x \left(\frac{1}{2}u^2\right) + ru^2\right) \\
&= \partial_t \int_{\mathbb{R}} \frac{1}{2}u^2 + \int_{\mathbb{R}} p(\partial_x u)^2 + \int_{\mathbb{R}} q\partial_x \left(\frac{1}{2}u^2\right) + \int_{\mathbb{R}} ru^2 \\
&= \partial_t \int_{\mathbb{R}} \frac{1}{2}u^2 + \int_{\mathbb{R}} p(\partial_x u)^2 + \underbrace{\lim_{N,N'\to\infty} \left(\frac{1}{2}u^2(N,t) - \frac{1}{2}u^2(N',t)\right)}_{0} + \int_{\mathbb{R}} ru^2 \\
&= \frac{1}{2}\partial_t \|u(\cdot,t)\|^2_{L^2(\mathbb{R})} + p\|\partial_x u\|^2 + r\|u(\cdot,t)\|^2.
\end{aligned}$$

Thus, this together with the Cauchy-Schwarz inequality implies

$$\frac{1}{2}\partial_t \|u(\cdot,t)\|^2_{L^2(\mathbb{R})} + r\|u(\cdot,t)\|^2 \le \int_{\mathbb{R}} uf \le \|u(\cdot,t)\|_{L^2(\mathbb{R})}\|f(\cdot,t)\|_{L^2(\mathbb{R})}.$$

By noting that $\frac{1}{2}\partial_t\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} = \|u(\cdot,t)\|_{L^2(\mathbb{R})}\partial_t\|u(\cdot,t)\|_{L^2(\mathbb{R})}$ (by Leibniz rule), we infer that

$$\partial_t\|u(\cdot,t)\|_{L^2(\mathbb{R})} + r\|u(\cdot,t)\|_{L^2(\mathbb{R})} \le \|f(\cdot,\mathbb{R})\|_{L^2(\mathbb{R})}.$$

Dropping the term $r\|u(\cdot,t)\|_{L^2(\mathbb{R})}$ and integrating in time, this gives the resulting L^2 a-priori error estimate

$$\|u(\cdot,T)\|_{L^2(\mathbb{R})} \le \|u_0\|_{L^2(\mathbb{R})} + \int_0^T \|f(\cdot,t)\|_{L^2(\mathbb{R})}dt.$$

□

Theorem 3.1.3. *Let $u(x,t)$ be a solution and $T > 0$. Then, we have the refined a-priori estimate*

$$\|u(\cdot,T)\|_{L^2(\mathbb{R})} \le e^{-rT}\|u_0\|_{L^2(\mathbb{R})} + \int_0^T e^{r(t-T)}\|f(\cdot,t)\|_{L^2(\mathbb{R})}dt.$$

Proof. The above computations give

$$\partial_t(e^{rt}\|u(\cdot,t)\|_{L^2(\mathbb{R})}) = e^{rt}\partial_t\|u(\cdot,t)\|_{L^2(\mathbb{R})} + re^{rt}\|u(\cdot,t)\|_{L^2(\mathbb{R})} \le e^{rt}\|f(\cdot,t)\|.$$

Integrating in time, we have the desired result. □

Theorem 3.1.4. *Let $u(x,t)$ be a solution and $T > 0$. Then, we have the a-priori estimate*

$$\frac{1}{2}\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} + \int_0^T \left(p\|\partial_x u(\cdot,t)\|^2_{L^2(\mathbb{R})} + r\|u(\cdot,t)\|^2_{L^2(\mathbb{R})}\right)dt \le \frac{1}{2}\|u_0\|^2_{L^2(\mathbb{R})} + \frac{1}{2r}\int_0^T \|f(\cdot,t)\|^2_{L^2(\mathbb{R})}dt.$$

Proof. By Young's inequality, we obtain that

$$\int_{\mathbb{R}} uf \le \|u(\cdot,t)\|_{L^2(\mathbb{R})}\|f(\cdot,t)\|_{L^2(\mathbb{R})} \le \frac{r}{2}\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} + \frac{1}{2r}\|f(\cdot,t)\|^2_{L^2(\mathbb{R})}.$$

Then we obtain that

$$\frac{1}{2}\partial_t\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} + p\|\partial_x u(\cdot,t)\|^2_{L^2(\mathbb{R})} + r\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} \le \frac{r}{2}\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} + \frac{1}{2r}\|f(\cdot,t)\|^2_{L^2(\mathbb{R})}.$$

It follows that

$$\frac{1}{2}\partial_t\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} + p\|\partial_x u(\cdot,t)\|^2_{L^2(\mathbb{R})} + r\|u(\cdot,t)\|^2_{L^2(\mathbb{R})} \le \frac{1}{2r}\|f(\cdot,t)\|^2_{L^2(\mathbb{R})}.$$

Integrating in time, we obtain the desired result. □

We are now ready to obtain stability using these error estimates.

Theorem 3.1.5. *Let u^1 and u^2 correspond to the data $(f^1, u_0^1), (f^2, u_0^2)$, respectively. Then,*

$$\|(u^1 - u^2)(\cdot, T)\|_{L^2(\mathbb{R})} \leq \|u_0^1 - u_0^2\|_{L^2(\mathbb{R})} + \int_0^T \|(f^1 - f^2)(\cdot, t)\|_{L^2(\mathbb{R})} dt.$$

Proof. Let $\phi = u^1 - u^2$. Then linearity implies that

$$\begin{aligned} \partial_t \phi - p\partial_{xx}\phi + r\phi &= f^1 - f^2, \quad x \in \mathbb{R}, t \in \mathbb{R}^+ \\ \phi(x, 0) &= u_0^1(x) - u_0^2(x) \\ \lim_{x \to \pm\infty} \phi(x, t) &= 0. \end{aligned}$$

□

We can apply the first a-priori estimate to this and obtain the stability bound.

We have discovered a notion of continuity for the solution operator. If the difference in data is small, then the difference in the corresponding solutions will be small as well. Moreover, if we consider a sequence $\{f^n, u_0^n\}$ such that $f^n \to f$ and $u_0^n \to u_0$ in $L^2(\mathbb{R})$, then $u^n \to u$ in $L^2(\mathbb{R})$. A similar argument holds by using the second estimate. Many more classes of stability bounds can be obtained by using other a-priori estimates.

Remark 3.1.6. The last estimate shows that a good candidate for a smoothness class where the existence of a solution could be established is the space composed of the functions $v : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ for which $t \mapsto \|v(\cdot, t)\|_{L^2(\mathbb{R})}$ is continuous and the following quantity is bounded:

$$\int_0^T (p\|\partial_x u\|_{L^2(\mathbb{R})}^2 + r\|u\|_{L^2(\mathbb{R})}^2) dt < \infty$$

for all $T > 0$. We define the space

$$X = \left\{ v : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R} \mid v(\cdot, t)_{L^2(\mathbb{R})} \in C(\mathbb{R}^+, \mathbb{R}),\ \int_0^T (p\|\partial_x u\|_{L^2(\mathbb{R})}^2 + r\|u\|_{L^2(\mathbb{R})}^2) dt < \infty,\ \forall\, T > 0 \right\}.$$

The estimate also shows that it is likely to construct a sufficient condition on the data (f, u_0) for the existence of a solution where $\|u_0\|_{L^2(\mathbb{R})}$ is bounded and $\int_0^T \|f(\cdot, t)\|_{L^2(\mathbb{R})} dt$ are bounded for all $T > 0$. Accordingly, we define,

$$Y = \left\{ (f, u_0) \mid \|u_0\|_{L^2(\mathbb{R})} < \infty,\ \int_0^T \|f(\cdot, t)\|_{L^2(\mathbb{R})}\| dt < \infty,\ \ \forall\, T > 0 \right\}.$$

Here, we equip X and Y with the natural norm and corresponding topology.

3.1.3 Uniqueness

Using stability, it is easy to obtain uniqueness.

Theorem 3.1.7. *Consider u^1 and u^2 as solutions to the PDE in the space X for the same data (f, u_0) in the normed space Y. Then, $u^1 = u^2$.*

Proof. The stability bound implies that for any $T > 0$,

$$\|(u^1 - u^2)(\cdot, T)\|_{L^2(\mathbb{R})} \leq \underbrace{\|u_0 - u_0\|_{L^2(\mathbb{R})}}_{0} + \int_0^T \underbrace{\|(f - f)(\cdot, t)\|_{L^2(\mathbb{R})}}_{0} dt = 0$$

and therefore $\|(u^1 - u^2)(\cdot, T)\|_{L^2(\mathbb{R})} = 0$. But this implies $u^1(\cdot, T) = u^2(\cdot, T)$ for all T and therefore $u^1 = u^2$. □

3.1.4 Existence

Proving existence of a solution in X with data in Y is quite technical. There are many methods of doing so. For example, one may approach the problem by using a Fourier transform in space. Another method that is closer to numerical analysis consists of constructing finite-dimensional approximations that are uniformly bounded in X and passing to the limit. This second method works particularly well when the space domain is a bounded interval $(-L, L)$. A general existence result for parabolic equations is known in the literature as Lions' theorem.

3.1.5 More on Boundary Conditions

We assume in this section that the model problem is set over the finite interval $(-L, L)$. In this case, many boundary conditions can be enforced. As we have seen above, deriving a priori estimates is essential to define a smoothness class where one can prove existence, uniqueness, and stability of a solution. All the arguments in- voked above using the energy method can be applied. The key point is the integration by parts in

$$\begin{aligned}-\int_{-L}^{L} pu\partial_{xx}u + \int_{-L}^{L} qu\partial_x u &= \int_{-L}^{L} p(\partial_x u)^2 + \int_{-L}^{L} q\partial_x \frac{1}{2}u^2 - pu\partial_x u\Big|_{-L}^{L} \\ &= \int_{-L}^{L} p(\partial_x u)^2 + (q\frac{1}{2}u^2 - pu\partial_x u)\Big|_{-L}^{L}\end{aligned}$$

Then, admissible boundary conditions are obtained by ensuring that the boundary terms appearing above produce non-negative terms. For instance, we could try to enforce

$$u(L)\left(q\frac{1}{2}u(L) - p\partial_x u(L)\right) \geq 0$$

$$u(-L)\left(q\frac{1}{2}u(-L) - p\partial_x u(-L)\right) \leq 0.$$

This can be achieved by enforcing Dirichlet boundary conditions: $u(L) = u(-L) = 0$. If $q \geq 0$, one can also enforce a Dirichlet boundary condition at $-L$ and a Neumann boundary condition at $+L$: $u(-L) = 0$, $\partial_x u(L) = 0$ and the other way around if $q \leq 0$. One can also enforce Robin boundary conditions $-\partial_x u(L) = Hu(L)$, and $\partial_x u(-L) = Hu(-L)$ where H is such that $H > -\frac{1}{2}q$.e

3.2 Lecture 2: Scalar Conservation Equations

3.2.1 Problem Statement

The goal of this lecture is to investigate the following nonlinear partial differential equation. We want to find $u : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ so that

$$\begin{aligned}\partial_t u + \partial_x(f(u)) &= 0, \quad x \in \mathbb{R}, x \in \mathbb{R}^+ \\ u(x,0) &= u_0(x), \quad x \in \mathbb{R} \\ \lim_{x\to\pm\infty}(u(x,t) - u_0(x)) &= 0.\end{aligned}$$

Here $f : \mathbb{R} \to \mathbb{R}$ is the **flux** and is a locally Lipschitz function, while u_0 is the initial data.

Example 3.2.1.

1. Linear tranport: $f(v) = \beta v$, $\beta \neq 0$.
2. Burgers equation: $f(v) = \frac{1}{2}v^2$.
3. Traffic flow equations: $f(p) = u_{\max}\rho(1 - \rho/\rho_{\max})$.
4. Buckley-Leverett equation: $f(v) = \frac{v^2}{v^2+(1-v)^2}$.

We again consider the problem of well-posedness, that is, obtaining existence, uniqueness, and stability.

3.2.2 Method of Characteristics

Let us assume that we have a unique solution $u(x,t)$ and let us assume that this solution is locally Lipschitz with respect to x and continuous with respect to t, at least over some time interval $t \in (0,T)$ The idea is to introduce a change a variable based on u.

Definition 3.2.2. *For $s \in \mathbb{R}$, the curve $\{(X(s,t),t) | t \geq 0\}$ in $\mathbb{R}\times[0,\infty)$ is called a* ***characteristic*** *where $X : \mathbb{R}\times\mathbb{R}^+ \to \mathbb{R}$ solves*

$$X_t(s,t) = f'(u(X,st),t), \quad s \in \mathbb{R}, t > 0$$
$$X(s,0) = s, \quad s \in \mathbb{R}.$$

Remark 3.2.3. Notice that, owing to the assumption we made on the solution u, the Cauchy-Lipchitz theorem (a.k.a. Picard-Lindelöf theorem) implies that $X(s,t)$ is well defined for all $s \in \mathbb{R}, t \in (0,T)$.

For the time being the situation looks desperate since $X(s,t)$ is defined by invoking u which is still unknown, but a little miracle will happen and will solve this conondrum. Consider a new functon $\phi : \mathbb{R}\times\mathbb{R}^+ \to \mathbb{R}$ defined by $\phi(s,t) = u(X(s,t),t)$. Then, by the chain rule

$$\phi + t(s,t) = u_x(X(s,t),t) \quad \underbrace{X_t(s,t)}_{f'(u(X(s,t)))} \quad + u_t(X(s,t),t)$$
$$= [u_x f'(u) + u_t](X(s,t),t) = 0$$

which implies that

$$\phi(s,t) = \phi(s,0) = u(X(s,0),0) = u_0(X(s,0)) = u_0(s).$$

Moreover, since we have that $X_t(s,t) = f'(u_0(s))$, then solving the ODE we obtain the implicit representation

$$u(X(s,t),t) = u_0(s), \quad X(s,t) = s + tf'(u_0(s)).$$

Thus, the characteristics are really just straight lines. Given $s \in \mathbb{R}$ and $t > 0$, the value of u at the location $X(s,t)$ and time t is $u_0(s)$. Obtaining an explicit representation of the solution to (1) using the methods of characteristics is in general nontrivial. This is done by expressing s as a function of x and t. Let $x \in \mathbb{R}$, $t > 0$. To find u, we must find $s \in \mathbb{R}$ such that

$$s + tf'(u_0(s)) = x.$$

This equatuon is nonlinear from the presence of of $f'(u_0(s))$. We can apply the Implicit Function Theorem to $G(s) = 0$ with $G(s) = s + tf'(u(s)) - x$. Let $f \in C^2(\mathbb{R})$. If $1 + tf''(u_0)\partial_x u_0(x) \neq 0$, then there is an $S(x,t)$ such that $G(S(x,t)) = 0$. With this function $S(x,t)$, we have

$$u(x,t) = u_0(S(x,t)).$$

In general, we have the following theorem:

Theorem 3.2.4. *Assume that $f \in C^2(\mathbb{R})$,$u_0 \in C^1(\mathbb{R})$, and $\inf_{\mathbb{R}}\{f''(u_0)u_0'\} > -\infty$ (essential lower bound). Then, the problem has a unique solution u over $t \in (0,T^*)$, where*

$$T^* = \begin{cases} \infty & \inf\{f''(u_0)u_0'\} \geq 0 \\ -\frac{1}{\inf\{f''(u_0)u_0'\}} & \{f''(u_0)u_0'\} < 0. \end{cases}$$

Let us assume that $u_0 \in C^1(\mathbb{R})$. If u_0, f are such that $1 + tf''(u_0(s))\partial_x u_0(s) \neq 0$, for all $s \in \mathbb{R}$ and $t > 0$, then $S(x,t)$ is always well-defined. In this case $T^* = \infty$. This above situation occurs when f is convex and u_0 is montonically increasing. The same conclusion holds if f is concave and u_0 is montonically decreasing.

Example 3.2.5. Consider the transport equation where $f(v) = \beta v$ for $\beta \neq 0$. Then $u_t + \beta u_x = 0$. The implicit representation gives

$$X(s,t) = s + tf''(u_0(s)) = s + \beta t.$$

Thus, for all $x \in \mathbb{R}$, $t > 0$,

$$u(x,t) = u(x + \beta t).$$

Example 3.2.6. Consider the Burgers' equation $f(v) = \frac{1}{2}v^2$. Then the PDE is given by $u_t + uu_x = 0$. We take the initial condition

$$u_0 = \begin{cases} 1, & x \le 0 \\ 1-x, & 0 < x < 1 \\ 0, & x \ge 1. \end{cases}$$

From the implicit representation,

$$X(s,t) = s + tf'(u_0(s)) = s + tu_0(s) = \begin{cases} s+t, & s \le 0 \\ s+t(1-s) & 0 < s < 1 \\ s, & s \ge 1. \end{cases}$$

Drawing the characteristics, we see that at $s = 1$ the solution becomes discontinuous. For $s \ge 1$ the characteristics are vertical lines while for $s < 1$, the characteristics are linear with slope 1. Thus, the characteristics will intersect for $s \ge 1$ and are traced back to two different points. Therefore, we see that $T^* = 1$ and that smoothness is lost in finite time. We refer to this as the solution developing a **shock**. Solving on a case by case basis, we obtain

$$S(x,t) = \begin{cases} x-t & x \le t \\ \frac{x-t}{1-t}, & t < x < 1 \\ x, & x \ge 1 \end{cases}$$

where $u(x,t) = u_0(S(x,t))$. Thus,

$$u(x,t) = \begin{cases} 1, & x \le t \\ 1 - \frac{x-t}{1-t}, & t < x < 1 \\ 0, & x \ge 1. \end{cases}$$

Theorem 3.2.7 (Rankine-Hugoniot Speed). *The speed of a shock is*

$$\frac{f(u_R) - f(u_L)}{u_R - u_L}$$

where u_R is u on the right of the shock and u_L is u on the left of the shock.

3.2.3 Weak Solutions

In order to make sense of solutions that are not $C^1(\mathbb{R})$, because either the initial data is not C^1 or smoothness is lost at some finite time T^*, we now introduce the notion of weak solutions. A weak formulation is obtained by testing the equation with smooth test functions that are compactly supported in $\mathbb{R} \times \mathbb{R}^+$, say $\phi \in C_0^1(\mathbb{R} \times \mathbb{R}^+)$.

Definition 3.2.8. *We say that $u \in L^\infty_{\text{loc}}(\mathbb{R} \times \mathbb{R}^+)$ is a weak solution if*

$$-\int_{\mathbb{R}^+} \int_{\mathbb{R}} (u\phi_t + f(u)\phi_x) dxdt - \int_{\mathbb{R}} u_0(x)\phi(x,0)dx = 0, \quad \forall\, \phi \in C_0^1(\mathbb{R} \times \mathbb{R}^+).$$

The problem with this definition is that there is no uniqeunss.

Example 3.2.9. Let $u_0(x) = \begin{cases} 0, & x \le 0 \\ 1, & x > 0 \end{cases}$ in Burger's equation. Indeed, we have that

$$X(s,t) = s + u_0 t = \begin{cases} s, & s \le 0 \\ s+t, s > 0 \end{cases}.$$

Then by observing the characteristics, we see that there is an empty region in which we have no information from characteristics. This produces two different weak solutions.

1. Solution 1: Shock. We place an artificial shock in the empty region that imposes an artifical barrier between the values 0 and 1. Everything to the right of the shock is 1 and everything to the left is 0. Using the formula for shock speed, we have that $x - \frac{1}{2}t$ is the shock line.
2. Solution 2: Rarefaction. We impose a boudary with $u_2(x) = x/t$ for $0 \leq x \leq t$ so that the solution changes smoothly from 0 to 1. This is physically valid.

Thus, we extend our weak solutions to a notion of entropy solutions.

Theorem 3.2.10. *For $f \in \mathrm{Lip}(\mathbb{R};\mathbb{R}), u_0 \in L^\infty(\mathbb{R})$, there is a unique entropy solution that is both a weak solution and satisfies*

$$-\int_{\mathbb{R}^+}\int_{\mathbb{R}}(\eta(u)\phi_+ + q(u)\phi_x)dxdt - \int_{\mathbb{R}}\eta(u_0)\phi(\cdot,0)dx \leq 0, \quad \forall\, \phi \in C_0^1(\mathbb{R}\times\mathbb{R}^+;\mathbb{R}^+)$$

and all entropy pairs (η, q). In other words, $\partial_t\eta(u) + \partial_x q(u) \leq 0$ in the sense of distributions.

3.3 Lecture 3: Wave Equation

3.3.1 Problem Statement

The goal of this lecture is to investigate the following nonlinear partial differential equation. We want to find $u : \mathbb{R}\times\mathbb{R}^+ \to \mathbb{R}$ so that

$$\begin{aligned} \partial_{tt} - c^2\partial_{xx}u &= 0, \quad x \in \mathbb{R}, x \in \mathbb{R}^+ \\ u(x,0) &= f(x), \quad x \in \mathbb{R} \\ u_t(x,0) &= g(x), \quad x \in \mathbb{R} \lim_{x\to\pm\infty}(u(x,t) - f(x)) &= 0. \end{aligned}$$

Here $f, g : \mathbb{R} \to \mathbb{R}$ are the initial data and c is the **wave speed**. This PDE is referred to as the **wave equation** and it is a hyperbolic problem. In this lecture we construct a solution to this problem using the Fourier transform technique.

3.3.2 Fourier Transform

Definition 3.3.1. *Let $f \in L^1(\mathbb{R})$. We define the Fourier transform of f, denoted $\mathcal{F}(f) : \mathbb{R} \mapsto \mathbb{C}$, such that*

$$\mathcal{F}(f)(\omega) = \frac{1}{2\pi}\int_{\mathbb{R}} f(x)e^{i\omega x}dx.$$

Indeed, this definition makes sense since

$$\left|\int_{\mathbb{R}} f(x)e^{i\omega x}dx\right| \leq \int_{\mathbb{R}} |f(x)||e^{i\omega x}|dx = \|f\|_{L^1(\mathbb{R})} < \infty.$$

Definition 3.3.2. *Let $f \in L^1(\mathbb{R})$. We define the inverse Fourier transform of f, denoted $\mathcal{F}^{-1}(f) : \mathbb{R} \mapsto \mathbb{C}$ such that*

$$\mathcal{F}^{-1}(f)(\omega) = \int_{\mathbb{R}} f(x)e^{i\omega x}dx.$$

Remark 3.3.3. Note that many authors will often swap these definitions. Also, many will change the constant from $1/2\pi$ and 1 to $1/\sqrt{2\pi}$ in both so that the Fourier transform is unitary. All of these are simply conventions.

Theorem 3.3.4. *Let $f \in L^1(\mathbb{R}) \cap C^1(\mathbb{R})$. Then, $\mathcal{F}^{-1}[\mathcal{F}(f)](x) = f(x)$ for all $x \in \mathbb{R}$. If f is discontinuous at x_0 but piecewise C^1, then*

$$\mathcal{F}^{-1}[\mathcal{F}(f)](x) = \frac{f(x_0^-) + f(x_0^+)}{2}.$$

Example 3.3.5. Here are examples of Fourier transform of some standard functions.

$$\mathcal{F}(e^{-\alpha|x|})(\omega) = \frac{1}{\pi}\frac{\alpha}{\omega^2+\alpha^2}, \quad \mathcal{F}\left(\frac{2}{x^2+\alpha^2}\right)(\omega) = e^{-\alpha|\omega|}.$$

$$\mathcal{F}(e^{-\alpha x^2})(\omega) = \frac{1}{\sqrt{4\pi\alpha}} e^{-\frac{\omega^2}{4\alpha}}.$$

$$\mathcal{F}(H(x)e^{-\alpha x})(\omega) = \frac{1}{2\pi}\frac{1}{\alpha - i\omega}.$$

Theorem 3.3.6. *Let $f \in L^1(\mathbb{R})$ and assume also that $\partial_x f \in L^1(\mathbb{R})$. Then,*

$$\mathcal{F}(\partial_x f)(\omega) = -i\omega\mathcal{F}(f)(\omega), \quad \forall\, \omega \in \mathbb{R}.$$

Moreover, if $f^{(n)} \in L^1(\mathbb{R})$, then

$$\mathcal{F}(f^{(n)})(\omega) = (-i\omega)^n \mathcal{F}(f)(\omega).$$

Remark 3.3.7. Let $f : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$. Assume that for all $t \in \mathbb{R}^+$, $f(\cdot, t) \in L^1(\mathbb{R})$ and $\partial_t f(\cot, t) \in L^1(\mathbb{R})$. Then,

$$\mathcal{F}(\partial_t f(\cdot, t))(\omega) = \partial_t \mathcal{F}(f(\cdot, t))(\omega).$$

Lemma 3.3.8. *Let $f \in L^1(\mathbb{R})$ and $\beta \in \mathbb{R}$. Then,*

$$\mathcal{F}(f(x - \beta))(\omega) = e^{i\beta\omega}\mathcal{F}(f)(\omega).$$

We now introduce the notion of convolution product.

Definition 3.3.9. *Let $f, g \in L^1(\mathbb{R})$. We define the function $f \star g$, called the convolution product of f and g, by*

$$(f \star g)(x) = \int_{\mathbb{R}} f(y)g(x - y)dy.$$

Lemma 3.3.10. *For all $f, g \in L^1(\mathbb{R})$, $f \star g = g \star f$.*

Theorem 3.3.11. *Let $f, g \in L^1(\mathbb{R})$. Then,*

$$\mathcal{F}(f \star g) = 2\pi\mathcal{F}(f)\mathcal{F}(g).$$

3.3.3 The d'Alembert Formula

Let us take the Fourier transform of the wave equation. We have

$$\mathcal{F}[\partial_{tt}u] - c^2\mathcal{F}[\partial_{xx}u] = 0$$

which implies from our results in the previous section that

$$\partial_{tt}\mathcal{F}[u] + c^2\omega^2\mathcal{F}[u] = 0.$$

This is an easy ODE to solve and we have that

$$\mathcal{F}[u](\omega, t) = A(\omega)e^{i\omega ct} + B(\omega)e^{-i\omega ct}$$

for constants $A(\omega), B(\omega)$. Fourier transforming the PDE data, we have $\mathcal{F}\sqcap(\omega, 0) = \mathcal{F}[f](\omega)$ and similarly for g. Thus, we obtain

$$A(\omega) + B(\omega) = \mathcal{F}[f](\omega)$$
$$i\omega c(A(\omega) - B(\omega)) = \mathcal{F}[g](\omega).$$

This implies

$$A(\omega) = \frac{1}{2}\mathcal{F}[f](\omega) + \frac{1}{2i\omega c}\mathcal{F}[g](\omega),$$
$$B(\omega) = \frac{1}{2}\mathcal{F}[f](\omega) - \frac{1}{2i\omega c}\mathcal{F}[g](\omega).$$

Thus,

$$\mathcal{F}[u](\omega, t) = \left(\frac{1}{2}\mathcal{F}[f](\omega) + \frac{1}{2i\omega c}\mathcal{F}[g](\omega)\right)e^{i\omega ct} + \left(\frac{1}{2}\mathcal{F}[f](\omega) - \frac{1}{2i\omega c}\mathcal{F}[g](\omega)\right)e^{-i\omega ct}.$$

Note from the shift lemma that

$$\mathcal{F}[f](\omega)e^{i\omega ct} + \mathcal{F}[f](\omega)e^{-i\omega ct} = \mathcal{F}[f(x-ct) + f(x+ct)](\omega).$$

Let us define $G(x) = \int_0^x g(\xi)d\xi$. Then $\partial_x G(x) = g(x)$ and $-i\omega\mathcal{F}[G](\omega) = \mathcal{F}[g](\omega)$. This shows that

$$\frac{1}{i\omega}\mathcal{F}(g)(\omega)e^{i\omega ct} - \frac{1}{i\omega}\mathcal{F}(g)(\omega)e^{-i\omega ct} = -\mathcal{F}[G](\omega)e^{i\omega ct} + \mathcal{F}[G](\omega)e^{-i\omega ct}$$

and by the shift lemma,

$$\frac{1}{i\omega}\mathcal{F}(g)(\omega)e^{i\omega ct} - \frac{1}{i\omega}\mathcal{F}(g)(\omega)e^{-i\omega ct} = -\mathcal{F}[G(x-ct) + G(x+ct)](\omega).$$

Putting everything together,

$$\begin{aligned}\mathcal{F}(u) &= \mathcal{F}\left(\frac{1}{2}(f(x-ct) + \frac{1}{2}f(x+ct)) + \frac{1}{2c}(G(x+ct) - G(x-ct))\right)\\ &= \mathcal{F}\left(\frac{1}{2}(f(x-ct) + \frac{1}{2}f(x+ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} g(\xi)d\xi\right).\end{aligned}$$

Taking the inverse Fourier transform, we have established the following result.

Theorem 3.3.12. *The unique weak solution to the wave equation is*

$$u(x,t) = \frac{1}{2}(f(x-ct) + f(x+ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} g(\xi)d\xi.$$

To convince ourselves that when f and g are smooth, this solution is unique, we can use an energy approach. We multiply by $\partial_t u$ so that

$$\begin{aligned}0 &= \int_{\mathbb{R}}[u_{tt}u_t - c^2 u_{xx}u_t]dx\\ &= \frac{d}{dt}\int_{\mathbb{R}}\frac{u_t^2}{2}dx - c^2\int_{\mathbb{R}} u_{xx}u_t dx\\ &= \frac{d}{dt}\int_{\mathbb{R}}\frac{u_t^2}{2}dx + c^2\int_{\mathbb{R}} u_x u_{tx} dx - \underbrace{c^2 u_x u_t\Big|_{-\infty}^{\infty}}_{0 \text{ by decay}}\\ &= \frac{d}{dt}\int_{\mathbb{R}}\frac{u_t^2}{2}dx + c^2\frac{d}{dt}\int_{\mathbb{R}}\frac{u_{x^2}}{2}dx\end{aligned}$$

so we obtain that

$$\frac{d}{dt}[\|u_t\|^2_{L^2(\mathbb{R})} + \|u_x\|^2_{L^2(\mathbb{R})}] = 0.$$

The quantity inside of the derivative is referred to energy and we see that

$$E(t) = E(0) = \int_{\mathbb{R}}[u_t(x,0)^2 + c^2 u_x(x,0)^2]dx = \int_{\mathbb{R}}[g(x)^2 + c^2 f(x)^2] = \|g\|^2_{L^2(\mathbb{R})} + c^2\|f\|^2_{L^2(\mathbb{R})}.$$

Now if u_1 and u_2 are solutions corresponding to the data (f,g), then the energy implies for $w = u_1 - u_2$, $\|w_t(\cdot,t)\|^2_{L^2(\mathbb{R})} = 0$ and $\|w_x(\cdot,t)\|^2_{L^2(\mathbb{R})} = 0$. This implies that w is constant in space and time. The initial condition $w(x,\cdot) = 0$ implies $w = 0$ identically.

4 Day 4: Finite Difference Methods

4.1 Lecture 1: Finite Differences Approximation

4.2 Lecture 2: Time-Domain Problems

4.3 Lecture 3: Time-Domain Problems Cont.

5 Day 5: Finite Element Methods

5.1 Lecture 1: Preliminaries

5.1.1 Motivation

Consider the following ODE:

$$-(p(x)u(x))' = f(x), \quad x \in (0,1)$$

where $p, f : (0,1) \to \mathbb{R}$ are given with $p(x) > 0$ for all $x \in (0,1)$ and $u : (0,1) \to \mathbb{R}$ an unknown function to be found.

Example 5.1.1. A typical example modeled by the above is the equilibrium temperature u of a rod represented by the interval $(0,1)$, given a heat conductivity p and a heat source f.

The ODE does not uniquely determine the solution u. In addition, we need to include boundary conditions. We shall focus on the Dirichlet boundary conditions

$$u(0) = \alpha, \quad u(1) = \beta,$$

where $\alpha, \beta \in \mathbb{R}$ are given.

Example 5.1.2. In the setting of the previous example, Dirichlet conditions impose a fixed temperature at the ends of the rod. Neumann conditions impose temperature fluxes at the end.

Definition 5.1.3. *Let $C^0[0,1]$ be the space of continuous functions on $[0,1]$, and, for $m \geq 1$, $C^[0,1]$ the space of functions f such that $f^m \in C^0[0,1]$.*

Remark 5.1.4. The ODE appears to require $p \in C^1(0,1)$, $f \in C^0[0,1]$, and $u \in C^2(0,1)$. However, the energy is given by

$$\frac{1}{2}\int_0^1 p'(x)|u'(x)|^2 dx = \int_0^[1]f(x)u(x)dx$$

and requires less regularity. What is the expected regularity of u? Using weak derivatives, we establish that not even $u \in C^0[0,1]$ is necessary for the system to have finite energy.

In view of the previous remark, is it possible to construct numerical schemes that do not require "smooth" data and solutions?

5.1.2 Weak Derivatives

Definition 5.1.5. *We say that a function $f : (0,1) \to \mathbb{R}$ is* ***square-integrable*** *in $(0,1)$ if it is integrable and*

$$\int_0^1 |f(x)|^2 dx < \infty.$$

The set of all such function is denoted $L^2(0,1)$, that is,

$$L^2(0,1) = \left\{ f : (0,1) \to \mathbb{R} \textit{ integrable} \mid \int_0^1 |f(x)|^2 dx < \infty \right\}.$$

It is a Hilbert space equipped with the inner product and norm,

$$(f,g)_{L^2(0,1)} = \int_0^1 f(x)g(x)dx, \quad (f,f) = \|f\|^{1/2}_{L^2(0,1)}.$$

Lemma 5.1.6. *Let $f, g \in L^2(0,1)$. Then,*

$$(f, g)_{L^2(0,1)} \leq \|f\|_{L^2(0,1)} \|g\|_{L^2(0,1)}.$$

Example 5.1.7. The set $L^2(0,1)$ contains discontinuous functions. For example, consider

$$f(x) = \begin{cases} -1 & 0 < x < \frac{1}{2}, \\ \pi & x = \frac{1}{2}, \\ 1 & \frac{1}{2} < x < 1. \end{cases}$$

Then,

$$\int_0^[1] f(x) dx = \int_0^{1/2} (-1) dx + \int_{1/2}^1 1 dx = 0$$

and

$$\int_0^1 |f(x)|^2 = \int_0^1 1 dx = 1 < \infty$$

so $f \in L^2(0,1)$. Also notice that sets of measure zero (i.e. single points) do not contribute to the integral.

Let $v \in C^1[0,1]$ and note that for all $w \in C^1[0,1]$ with $w(0) = w(1) = 0$, integration by parts produces

$$\int_0^1 v'(x) w(x) = -\int_0^1 v(x) w'(x).$$

More compactly,

$$(v', w)_{L^2(0,1)} = -(v, w')_{L^2(0,1)}.$$

We ofte write $C_0^1[0,1]$ for continuously differentiable functions that are zero on the boundary. This justifies the following definition.

Definition 5.1.8. *Let $v \in L^2(0,1)$. We say that v has a* ***weak derivative*** *in $L^2(0,1)$ if there exists $\phi \in L^2(0,1)$ such that*

$$(\phi, w)_{L^2(0,1)} = -(v, w')_{L^2(0,1)}, \quad \forall\, w \in C_0^1[0,1].$$

In this case, we write $\phi = v'$.

We accept the following facts about weak derivatives:

- Changing the value of a function at one point does not change its weak derivative;
- If a weak derivative exists, it must be unique (up to the value at a finite number of points), and so generates an equivalence class.

The uniqueness of weak derivative implies that if $v \in C^1[0,1]$ then the standard derivative is also the weak derivative.

Example 5.1.9. Consider the function

$$v(x) = \begin{cases} 2x & 0 < x \leq \frac{1}{2}, \\ 2 - 2x & \frac{1}{2} < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

This function is not in $C^1[0,1]$. However, for any $w \in C_0^1[0,1]$, we have

$$\begin{aligned}
\int_0^1 v(x)w'(x) &= 2\int_0^{1/2} xw'(x)dx + 2\int_{1/2}^1 w'(x) - 2\int_{1/2}^1 xw'(x)dx \\
&= 2\left(-\int_0^{1/2} w(x)dx + xw(x)\bigg|_0^{1/2}\right) + 2w(1) - 2w(1/2) - 2\left(-\int_{1/2}^1 w(x)dx + xw(x)\bigg|_{1/2}^1\right) \\
&= -2\left(\int_0^1 w(x)dx - \int_{1/2}^1 w(x)dx\right) + 2\frac{1}{2}w(1/2) - 2w(1/2) - 2(w(1) - 1/2w(1/2)) \\
&= -2\left(\int_0^1 w(x)dx - \int_{1/2}^1 w(x)dx\right) \\
&= -\left(\int_0^{1/2} (2)w(x)dx + \int_{1/2}^1 (-2)w(x)dx\right).
\end{aligned}$$

Therefore, we identify

$$\phi(x) = \begin{cases} 2 & 0 < x \le \frac{1}{2} \\ -2 & \frac{1}{2} < x < 1 \end{cases}$$

so that we have

$$(v, w')_{L^2(0,1)} = -(\phi, w), \quad \forall\, w \in C_0^1[0,1].$$

Notice that $\phi \in L^2(0,1)$. Therefore, ϕ is the weak derivative of v. Notice that v' does not have a weak derivative in $L^2(0,1)$. Indeed, for all $w \in C_0^1[0,1]$,

$$\int_0^1 \phi(x)v'(x)dx = \int_0^{1/2} 2v'(x) - \int_{1/2}^1 2v'(x) = 4v(1/2)$$

and there is no $L^2(0,1)$ function ψ such that

$$4v(1/2) = -\int_0^1 \psi(x)v(x)dx.$$

In fact, $\psi(x) = -4\delta_{1/2}(x) \notin L^2(0,1)$, where $\delta_{1/2}(x)$ is the Dirac measure at $1/2$.

Definition 5.1.10. *We denote by $H^1(0,1)$ the* ***Sobolev space*** *of $L^2(0,1)$ functions having a weak derivative in $L^2(0,1)$, i.e.,*

$$H^1(0,1) = \{v \in L^2(0,1) \mid v' \in L^2(0,1)\}.$$

It is a Hilbert space with the inner product and norm,

$$(f,g)^2_{H^1(0,1)} = (f,g)^2_{L^2(0,1)} + (f',g')^2_{L^2(0,1)}, \quad \|f\|^2_{H^1(0,1)} = (f,f)^{1/2}_{H^1(0,1)}.$$

Like $L^2(0,1)$, the set $H^1(0,1)$ consists of equivalence classes of functions from their pointwise invariance. We will accept that for every $f \in H^1(0,1)$, there is an extension $\tilde{f} \in C^0[0,1]$, that is, $f = \tilde{f}$ almost everywhere. From now on, when we write $f \in H^1(0,1)$, we mean the continuous representation $\tilde{f}$ so that pointwise values of f are well-defined.

Lemma 5.1.11. *For $v, w \in H^1(0,1)$, it holds that*

$$\int_0^1 v'(x)w(x)dx = -\int_0^1 v(x)w'(x) + v'(x)w(x)\bigg|_0^1.$$

Definition 5.1.12. *The set of functions*

$$H_0^1(0,1) = \{v \in H^1(0,1) \mid v(0) = v(1) = 0\}$$

is the subset of $H^1(0,1)$ consisting of functions vanishing at 0 and 1.

Note that if v is smooth and satisfies $v(0) = 0$, then

$$v(x)^2 - v(0)^2 = \int_0^x (v(s)^2)' ds = 2\int_0^x v(s)v'(s)ds.$$

Therefore,

$$v(x)^2 \leq 2\int_0^1 |v(s)v'(s)|ds \leq 2\|v\|_{L^2(0,1)}\|v'\|_{L^2(0,1)}.$$

After integrating from 0 to 1, we deduce that

$$\|v\|^2_{L^2(0,1)} \leq 2\|v\|_{L^2(0,1)}\|v'\|_{L^2(0,1)}$$

or equivalently,

$$\|v\|^2_{L^2(0,1)} \leq 2\|v'(0,1).$$

This estimate is known as the **Poincare inequality** and is more generally true for functions in $H^1_0(0,1)$.

Lemma 5.1.13. *For $v \in H^1_0(0,1)$, there exists $C > 0$ such that*

$$\|v\|_{L^2(0,1)} \leq C\|v'\|_{L^2(0,1)}.$$

5.1.3 Weak Formulation

We return to the problem of finding $u \in C^2[0,1]$ satisfying

$$-(p(x)u'(x))' = f(x), \quad x \in (0,1), \quad u(0) = u(1) = 0.$$

We assume that $p \in L^2(0,1)$ is such that $0 < P_{\min} \leq p(x) \leq P_{\max}$ a.e. for some $0 < P_{\min} \leq P_{\max} < \infty$ and that $f \in L^2(0,1)$. Notice that in particular f is not necessarily continuous, which therefore requires us to give a different meaning to the ODE.

Remark 5.1.14. For the case with general Dirichlet boundary conditions $u(0) = \alpha$, $u(1) = \beta$, we set $u_0 = \alpha+(\beta-\alpha)x$ so that $\tilde{u} = u - u_0$, satisfies the ODE with zero boundary conditions with $f(x)$ replaced by $f(x) + (\beta - \alpha)p'(x)$.

For now, we assume $u \in C^2[0,1]$ and multiply the ODE by $v \in H^1_0(0,1)$ and integrate by parts

$$\int_0^1 p(x)u'(x)v'(x)dx - p(x)u'(x)v(x)\Big|_0^1 = \int_0^1 f(x)v(x)dx.$$

Because $v(0) = v(1) = 0$,

$$\int_0^1 p(x)u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx, \quad \forall\, v \in H^1_0(0,1).$$

Notice that from Cauchy-Schwarz and the assumptions on p and f, we have

$$\left|\int_0^1 p(x)u'(x)v'(x)dx\right| \leq P_{\max}\|u\|_{H^1(0,1)}\|v\|_{H^1(0,1)} < \infty$$

and

$$\left|\int_0^1 f(x)v(x)dx\right| \leq \|f\|_{L^2(0,1)}\|v\|_{H^1(0,1)} < \infty.$$

This justifies the following definition.

Definition 5.1.15. *The **weak formulation** of the ODE is to find $u \in H^1_0(0,1)$ such that*

$$\int_0^1 p(x)u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx, \quad \forall\, v \in H^1_0(0,1).$$

*This $u \in H^1_0(0,1)$ is said to be a **weak solution** to the ODE.*

The next result state the existence and uniqueness of weak solutions to the ODE.

Lemma 5.1.16 (Lax-Milgram). *The weak formulation has a unique solution $u \in H^1_0(0,1)$. It satisfies the stability estimate*

$$\|u\|_{H^1(0,1)} \leq 2P^{-1}_{\min}\|f\|_{L^2(0,1)}.$$

5.2 Lecture 2: Finite Elements

5.2.1 Discretization

Our goal is to replace our infinite dimension space $H_0^1(0,1)$ in the weak formulation by a finite dimensional approximation space to compute a solution. Let

$$0 \le x_0 < x_1 < \cdots < x_{N+1} = 1$$

be an partition of $[0,1]$. For each $i = 1, \dots, N$, we define the "hat" function

$$\phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} & x \in [x_{i-1}, x_i], \\ \frac{x_{i+1} - x}{x_{i+1} - x_i} & x \in [x_i, x_{i+1}], \\ 0 & \text{otherwise.} \end{cases}$$

Each ϕ_i is piecewise linear and therefore in $H_0^1(0,1)$. Moreover, $\phi_i(x_j) = \delta_{ij}$ which implies that $\{\phi_i\}_{i=1}^N$ are linerly independent. To see this, assume that for some $\{\alpha_i\}_{i=1}^N \in \mathbb{R}$ there holds

$$\sum_{i=1}^N \alpha_i \phi_i(x) = 0, \quad \forall\, x \in [0,1].$$

Then, for $j \in [N]$, we have

$$0 = \sum_{i=1}^N \alpha_i \phi_i(x_j) = \sum_{i=1}^N \alpha_i \delta_{ij} = \alpha_j.$$

Therefore, this implies $\alpha_j = 0$ for all j and $\{\phi_i\}$ are linearly independent. Let $V_N = \mathrm{span}(\phi_1, \dots, \phi_N) \subset H_0^1(0,1)$, i.e.,

$$V_N = \left\{ \sum_{i=1}^N \alpha_i \phi_i(x) \mid \alpha_i \in \mathbb{R},\ i = 1, \dots, N \right\}.$$

From linear independence, V_N has dimension N. Define

$$W_N = \left\{ v \in C^0[0,1] \mid v(0) = v(1) = 0,\ v|_{[x_i, x_{i+1}]} \text{ is linear},\ i = 0, \dots, N \right\}.$$

Then, $V_N \subset W_N$ because each $\phi_i \in W_N$. Moreeover, any $w \in W_N$ can be written as

$$w(x) = \sum_{i=1}^N w(x_i)\phi_i(x)$$

which implies $w \in V_N$. Thus, $W_N \subset V_N$ and consequently $V_N = W_N$.

5.2.2 Linear System Formulation

We replace $H_0^1(0,1)$ by the finite dimensional subspace V_N.

Definition 5.2.1. *The discrete weak formulation reads: Find $u_N \in V_N$ such that*

$$\int_0^1 p(x) u_N'(x) v_N'(x) dx = \int_0^1 f(x) v_N(x) dx, \quad v_N \in V_N.$$

As we shall see, a unique solution $u_N \in V_N$ exists to the above problem. This u_N is called the finite element solution. Notice that since the discrete weak formulation holds for all $v_N \in V_N$, it also holds that

$$\int_0^1 p(x) u_N'(x) \phi_j(x) dx = \int_0^1 f(x) \phi_j(x) dx, \quad j = 1, \dots, N.$$

Since $u_N \in V_N$, we have that

$$u_N(x) = \sum_{i=1}^{N} U_i \phi_i(x)$$

for some set of coefficients $\{U_i\} \in \mathbb{R}$. Using this ansatz, we obtain that u_N is the finite element solution if and only if

$$\sum_{i=1}^{N} U_i \int_0^1 p(x)\phi_i'(x)\phi_j'(x)dx = \int_0^1 f(x)\phi_j(x)dx, \quad j = 1, \ldots, N.$$

Define the **stiffness matrix**

$$A = (a_{ij})_{i,j=1}^N, \quad a_{ij} = \int_0^1 p(x)\phi_i'(x)\phi_j'(x)dx$$

and the vectors

$$F = (F_j)_{j=1}^N, \quad F_j = \int_0^1 f(x)\phi_j(x), \quad U = (U_j)_{j=1}^N.$$

With this notation, u_N is the finite element solution if and only if its coefficients satisfy the linear system

$$AU = F.$$

Remark 5.2.2. Notice that this system is sparse. Indeed, when $|i-j| > 1$, $a_{ij} = 1$ so only the three primary diagonals are populated. The use of sparse matrices is critical because it requires (approximately) the storage of $3N$ doubles instead of N^2 doubles. Recall that 1 double is 8 bytes or 8×10^{-7} MB. For $N = 106$ the use of a sparse matrix needs 2.4 MB while a full matrix requires 800 GB. Moreover, Gaussian elimination (not used in practice) requires $O(N^3)$ operations for a full matrix. For an n-diagonal matrix, only nN operations are needed.

Remark 5.2.3. When $p(x) = 1$ and $x_i = ih$ with $h = 1/(N+1)$, A is the finite difference matrix up to a scaling factor.

5.3 Lecture 3: Well-posedness

5.3.1 Existence and Uniqueness

To show that there is a unique vector $U \in \mathbb{R}^N$ satisfying the linear system $AU = F$, we need to show that A is invertible. We show that $\text{Ker}(A) = \{0\}$. Assume that $AV = 0$ for some $V \in \mathbb{R}^N$. Therefore,

$$V^T AV = 0 \implies \sum_{i,j=1}^{N} V_i a_{ij} V_j = 0.$$

From the definition of a_{ij}, we find that

$$\int_0^1 p(x)|v_N'(x)|^2 dx = 0$$

where $v_N = \sum_{i=1}^N V_i\phi_i(x)$. Taking advantage of the assumption on p, we deduce that

$$\|v_N'\|_{L^2(0,1)} = 0 \implies v_N = 0.$$

From linear independence, we obtain that $V = 0$ and A is therefore injective. But since A is finite dimensional, it is surjective and therefore invertible.

5.3.2 Stability

Taking $v_N = u_N$, we find that

$$\int_0^1 p(x)|u_N'(x)|^2 dx = \int_0^1 f(x)u_N(x)dx.$$

From the Cauchy Schwarz inequality and the assumption on p,

$$P_{\min}\|u_N'\|^2_{L^2(0,1)} \leq \|f\|_{L^2(0,1)}\|u_N\|_{L^2(0,1)} \leq \|f\|_{L^2(0,1)}\|u_N\|_{H^1(0,1)}$$

or

$$\|u_N'\|^2_{L^2(0,1)} \leq P_{\min}^{-1}\|f\|_{L^2(0,1)}\|u_N\|_{H^1(0,1)}.$$

The Poincare inequality implies that there exists $C > 0$ such that

$$\|u_N\|_{L^2(0,1)} \leq \|u_N'\|_{L^2(0,1)}$$

and so

$$\|u_N\|_{H^1(0,1)} \leq CP_{\min}^{-1}\|f\|_{L^2(0,1)}.$$

Since in our case $C = 2$, we see that this is identical to the Lax-Milgram lemma

$$\|u\|_{H^1(0,1)} \leq 2P_{\min}^{-1}\|f\|_{L^2(0,1)}.$$

In particular, for two solutions u_N^1, u_N^2 corresponding to f^1, f^2, we have

$$\|u_N^1 - u_N^2\|_{H^1(0,1)} \leq 2P_{\mathrm{mi}}^{-1}\|f^- f^2\|_{L^2(0,1)}$$

and stability.

5.3.3 Convergence

For any $v_N \in V_N$, the weak solution satisfies

$$\int_0^1 p(x)u'(x)v_N'(x)dx = \int_0^1 f(x)v_N(x)dx$$

and the finite element solution satisfies

$$\int_0^1 p(x)u_N'(x)v_N'(x)dx = \int_0^1 f(x)v_N(x)dx.$$

By subtracting, we see that

$$\int_0^1 p(x)(u'(x) - u_N'(x))v_N'(x)dx = 0, \quad \forall\, v_N \in V_N.$$

This is the **Galerkin orthogonality**. The orthogonality refers to the fact that the error in the weak derivative of the finite element solution is in the orthogonal complement of V_N. In view of this, we compute

$$\begin{aligned}
\|u' - u_N'\|^2_{L^2(0,1)} &\leq \frac{1}{P_{\min}}\int_0^1 p(x)|u'(x) - u_N'(x)|^2 dx \\
&= \frac{1}{P_{\min}}\int_0^1 p(x)(u'(x) - u_N'(x))(u'(x) - u_N'(x))dx \\
&= \frac{1}{P_{\min}}\int_0^1 p(x)(u'(x) - u_N'(x))(u'(x) - v_N'(x))dx \\
&\leq \frac{P_{\max}}{P_{\min}}\|u' - u_N'\|^2_{L^2(0,1)}\|u' - v_N'\|^2_{L^2(0,1)}
\end{aligned}$$

and so

$$\|u' - u_N'\|_{L^2(0,1)} \leq \frac{P_{\max}}{P_{\min}} \min_{v_N \in V_N} \|u' - v_N'\|^2_{L^2(0,1)}.$$

This is known as the **best approximation property** as it says that the finite element solution is the best approximation in the chosen space V_N.

We will now show convergence. For now, assume that $u \in H^2(0,1)$. Define the linear interpolant of u

$$I_N u(x) = \sum_{i=1}^{N} u(x_i)\phi_i(x).$$

Notice that $u(x_i) = I_N u(x_i)$ for $i = 0, \dots, N+1$. Define $e(x) = u(x) - I_N u(x)$ so that $e(x_i) = 0$ for all $i = 0, \dots, N+1$. We apply Rolle's theorem to guarantee the existence of $\xi_j \in (x_j, x_{j+1})$ for $j = 0, \dots, N$ such that $e'(\xi_j) = 0$. For $x \in (x_j, x_{j+1})$, the Fundamental Theorem of Calculus implies

$$e'(x) = \int_{\xi_j}^{x} e''(s)ds = \int_{\xi_j}^{x} u''(s)ds.$$

From Cauchy-Schwarz,

$$e'(x)^2 \le |x - \xi_j| \int_{\xi_j}^{x} (u''(s))^2 ds \le |x_{j+1} - x_j| \int_{x_j}^{x_{j+1}} (u''(s))^2 ds$$

and by integrating,

$$\int_{x_j}^{x_{j+1}} e'(x)^2 \le \max_{j=0,\dots,N} |x_{j+1} - x_j|^2 \int_{x_j}^{x_{j+1}} (u''(s))^2 ds.$$

After summing over $j = 0, \dots, N+1$,

$$\|u' - (I_N u)'\|_{L^2(0,1)} \le \max_{j=0,\dots,N} |x_{j+1} - x_|| \|u''\|_{L^2(0,1)}.$$

Returning to $\|u' - u_N'\|_{L^2(0,1)}$, the best approximation property yields

$$\|u' - u_N'\|_{L^2(0,1)} \le \frac{P_{\max}}{P_{\min}} \|u''\|_{L^2(0,1)} \max_{j=0,\dots,N} |x_{j+1} - x_j|.$$

The right side tends to 0 whenever $\max_{j=0,\dots,N} |x_{j+1} - x_j| \to 0$. When the subdivision is uniform, that is $x_i = i/(N+1)$, we have

$$\|u' - u_N'\|_{L^2(0,1)} \le \frac{P_{\max}}{P_{\min}} \|u''\|_{L^2(0,1)} \frac{1}{N+1} \to 0$$

as $N \to \infty$.

6 Project

6.1 Preliminaries

Let $D \subset \mathbb{R}^3$. We care about studying the radiation $\Psi : D \times \mathbb{S}^2 \to \mathbb{R}$, $(x, \Omega) \mapsto \Psi(x, \Omega)$. In particular, we want to find Ψ such that

$$\Omega \cdot \nabla_X \Psi(x, \Omega) + \sigma^t \Psi(x, \Omega) = \frac{\Sigma^s}{|\mathbb{S}^2|} \int_{\mathbb{S}^2} \Psi(x, \Omega') d\Omega' + q(x), \quad \text{in } \Omega \times \mathbb{S}^2.$$

$$\Psi(x, \Omega) = \alpha^{\partial}(x), \quad \text{on } \{x \in \partial D, \ \Omega \in \mathbb{S}^2 \mid n_x \cdot \Omega < 0\}.$$

Here, σ^t is the total cross section, σ^s is the scattering cross section, and $\sigma^a = \sigma^t - \sigma^s$ the absorption cross section. The source q occurs from the physical nature of the problem, for example black-body radiation, in which the problem becomes couple with additional PDE constraints on q. We say that the Neumann condition is isotropic if there is no dependence on Ω. These problems are relevant to study for neutron scattering and nuclear fusion.

To numerically obtain a solution, we use the **discrete ordinates method**. Let us establish a quadrature rule over $\mathbb{S}^2$ with weights $\{w_l\}_{l \in \mathcal{L}}$ such that $\sum_l w_l = 1$. Then, we want to find $\{\psi^l\}_{l \in \mathcal{L}}$ such that

$$\Omega_l \cdot \nabla \psi^l + \sigma^t(x) \psi^l = \sigma^s(x) \sum_k \omega_k \psi^k + q(x), \quad \forall \, l \in \mathbb{L}$$

where $0 \leq \sigma^s \leq \sigma^t$ for all $x \in D$. We will assume that σ^s, σ^t, q are piecewise constant on each cell in our mesh.

For now, let us simplify this problem to the interval $[a, b]$. Fix $\mu \neq 0$. If $\mu > 0$, we have an inflow problem and prescribe a left boundary. If $\mu < 0$, we have an outflow problem and prescribe a right boundary. We want to find $u : [a, b] \to \mathbb{R}$ such that

$$\mu u' + \sigma^t u = q$$

$$u(a) = \alpha.$$

Multiplying by a suitable test function v and integrating, we have

$$\int_a^b \mu u' v + \sigma^t u v = \int_a^b q v dx.$$

Hence, the discrete weak formulation is to find $u_h \in V_h$ such that

$$\int_a^b \mu u_h v_h + \sigma^t u_h v_h = \int_a^b q v_h dx, \quad \forall \, v_h \in V_h.$$

Taking $u_h = \sum_{i=0}^{N-1} u_i \phi_i$ and testing with $v_h = \phi_j$, we have

$$\sum_{i=1}^{n} u_i \int_a^b (\mu \phi_i' \phi_j + \sigma^t \phi_i \phi_j) dx = \int_a^b q \phi_j dx, \quad i = 0, \ldots, N-1.$$

Define

$$(T)_{ij} = \int_a^b (\phi_i' \phi_j + \sigma^t \phi_i \phi_j), \quad b_i = \int_a^b q \phi_j dx.$$

This is an $N-1 \times N-1$ system. To impose the boundary condition, we artifically impose it into the linear system through an additional row. From here, we simply solve $Tu = b$ for the coefficient vector u.

Solving this system, we see that we have stability issues from spurious oscillations in the solution. This occurs from applying a naive Galerkin scheme to first order PDE. To fix this, we utilize the Streamline-Upwind Petrov-Galerkin (SUPG) scheme. We instead consider the weak formulation: Find $u \in V$ such that

$$\int_a^b (\mu u' + \sigma^t u)(v + \tau \mu v') dx = \int_a^b q(v + \tau \mu v') dx, \quad \forall v \in V.$$

The upwind approach comes from modifying the test function term with its derivative. Moreover, the method is Petrov-Galerkin as the test functions lie in a different space V. The coefficient τ is defined by

$$\tau = \xi \max(|\mu| h^{-1}, \sigma^t)^{-1}$$

where ξ is a tuning parameter. Generally, we take $\xi \approx 1$ and $\tau = \tau_k$ that varies over each cell.

We now have a code that can solver transport for one fixed problem. Suppose now that we project the sphere onto $[-1, 1]$ with a quadrature set $\mathcal{L}$ with weights

$$\sum_{l \in \mathcal{L}} w_l = 1, \quad \sum_l \omega_l f(\mu_l) \approx \frac{1}{2} \int_{[-1,1]} f(\mu) d\mu.$$

We want to find $\{\Psi_h^l\}_{l \in \mathcal{L}} \in V_h^{\mathcal{L}}$ (one for every angle) such that

$$t_h^l(\Psi_h^l, \phi) + s_h^l(\Psi_h^l, \phi) + b_h^l(\Psi_h^l, \phi) = (\sigma^s \sum_l w_l \Psi_h^l, \phi + \tau \mu \phi') + (q, \phi + \tau \mu^l \phi') + b_h^l(\Psi_h^l, \phi), \quad \forall \, \phi \in V_h$$

Define $\Psi_h^{l,*} \in V_h$ that solves

$$t_h^l(\Psi_h^{l,*}, \phi) + s_h^l(\Psi_h^{l,*}, \phi) + b_h^l(\Psi_h^{l,*}, \phi) = (q, \phi + \tau \mu^l \phi') + b_h^l(\Psi_h^{l,*}, \phi), \quad \forall \, \phi \in V_h.$$

In other words, it solves

$$\mu^l u' + \sigma^t u = q$$

along with the boundary conditions. Define $\Psi_h^{l,0} : V_h \to V_h$ that solves $\mu^l u' + \sigma^t u = \sigma^s = \varphi$, i.e,

$$t_h^l(\psi_h^{l,0}(\varphi), \Phi) + s_h^0(\Psi_h^{l,0}(\varphi), \phi) = (\sigma^2 \varphi, \phi + \tau \mu^l \phi').$$

The desired solution

$$\Gamma_h^0(\phi) = (\Psi_h^{1,0}, \Psi_h^{2,0}, \dots, \Psi^{|\mathcal{L}|,0}),$$
$$\Gamma_h^*(\phi) = (\Psi_h^{1,*}, \Psi_h^{2,*}, \dots, \Psi_h^{|\mathcal{L},*})$$
$$\phi = \sum_l w_l(\Psi_h^{l,0} + \Psi_h^{l,*}) = \overline{\Psi_h^{l,0}(\phi)} + \overline{\Psi_h^{l,*}}.$$

To summarize, we

- Compute $\overline{\Psi_h^{l,*}}$ by applying our previous method for every angle.
- We make an initial guess for $\phi^{(0)}$.
- Then we compute
$$\phi^{(n+1)} = \overline{\Psi n^{l,0}(\phi^{(n)})} + \overline{\Psi^{l,*}}$$
- Finally, we construct the actual intensity
$$\Psi_h^l = \Psi^{l,0}(\phi_{\text{soln}}) + \Psi^{l,*}.$$

Going back to our original problem:

$$\Omega \cdot \nabla_x \Psi + \sigma^t(x)\Psi = \frac{\sigma^t(x)}{|\mathbb{S}^2|} \int_{\mathbb{S}^2} \psi(x, \Omega') d\Omega' + q(x)$$

where $\Psi = \alpha^\partial$ on the inflow boundary. Let $\epsilon > 0$. We can non-dimensionalize this PDE into the form

$$\Omega \cdot \nabla_x \Psi + \sigma \left(\epsilon + \frac{1}{\epsilon}\right) \Psi = \frac{\sigma}{\epsilon |\mathbb{S}^2|} \int_{\mathbb{S}^2} \psi(x, \Omega') d\Omega' + \epsilon \overline{q}(x)$$

where $\Psi = \epsilon \overline{alpha}^\partial$ on the inflow boundary. We have the uniform limit

$$\Psi = \Psi^{(0)} + \epsilon \Psi^{(1)} + \epsilon^2 \Psi^{(2)} + \cdots$$

where $\Psi^{(0)}$ is isotropic (no dependence on Ω) and satisfies the PDE

$$-\nabla \left(\frac{1}{3\sigma^s} \nabla \Psi^{(0)}\right) + \sigma^a \Psi^{(s)} = q$$

where $\Psi^{(0)} = \alpha^\partial$ on ∂D.