

PDE & Data Control Seminar

after slide 7
part 4:

weak formulation

$$\textcircled{*} \quad \begin{aligned} -\Delta u &= f && \text{in } \Omega \\ u &= 0 && \text{on } \Gamma \end{aligned} \quad \left. \begin{array}{l} \text{elliptic boundary} \\ \text{value problem} \end{array} \right\}$$

$f \in L^2(\Omega)$ given \rightarrow can be very irregular
 \rightarrow no classical solution $y \in C^2(\bar{\Omega}) \cap C(\bar{\Omega})$
 \rightarrow seek weak solution $y \in H_0^1(\Omega)$

Assume that Ω is bounded,
multiply by arbitrary test function $v \in C_c^\infty(\Omega)$ and
integrate over Ω :

$$-\int_{\Omega} v \Delta u \, dx = \int_{\Omega} f v \, dx$$

integration
by parts

$$\Rightarrow -\int_{\substack{\Gamma \\ =0 \text{ on } \Gamma}} v \partial_{\nu} u \, ds + \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx$$

$$\Rightarrow \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \textcircled{*}^2$$

$\partial_{\nu} u$: normal derivative of u , i.e., the directional derivative of u in the direction of the outward unit normal ν to Γ

$C_0^\infty(\Omega)$ is dense in $H_0^1(\Omega)$

for fixed y everything depends continuously on $v \in H_0^1(\Omega)$

→ equation holds for all $v \in H_0^1(\Omega)$

It can be proven that any sufficiently smooth $u \in H_0^1(\Omega)$ satisfying $\textcircled{*}^2$ solves $\textcircled{*}^1$.

Definition

We call $u \in H_0^1(\Omega)$ a weak solution to $\textcircled{*}^1$ if it satisfies the so-called weak formulation

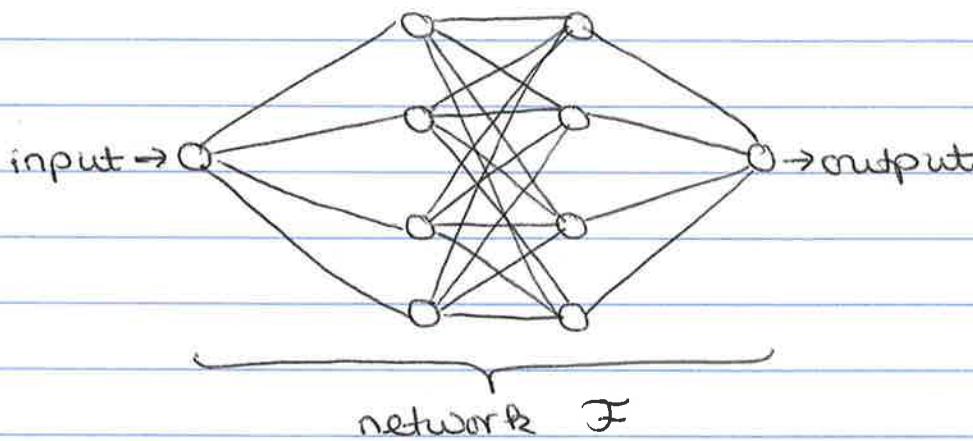
$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H_0^1(\Omega).$$

after slides

Example 3: Machine Learning

- Aim:
- Classification (Is this email spam? Yes or No)
 - Regression (predict a numerical value)
 - Transcription (transcribe into discrete textual form)
 - Machine translation (language translation)
 - ...

given data set: $\{u^{(i)}, S(u^{(i)})\}_{i=1}^N$



$$F = f_2 \circ f_1 \circ f_0$$

simple feed forward network

$$y^{[l]} = f_{l+1}(y^{[l-1]}, \Theta^{[l-1]}) \quad , l=1, \dots, L$$

$$y^{[0]} = u, \quad \Theta^{[l-1]} = (W^{[l-1]}, b^{[l-1]})^T \quad (\text{weights and biases})$$

→ control the system (learn Θ) with respect
to a control goal

control goal:

$$y^{[L](i)} \approx S(u^{(i)}) \quad \text{for all } i=1, \dots, N$$

$$\min_{y, \theta} J(y, \theta) = \frac{1}{2N} \sum_{i=1}^N \|y^{[L](i)} - S(u^{(i)})\|_2^2$$

mean squared error

- choice of loss function depends on data
- another example: Cross entropy

problem:

$$\min_{y, \theta} J(y, \theta) \quad \text{such that} \quad y^{[L](i)} = \mathcal{F}(u^{(i)}, \theta)$$

y : feature vectors \leftrightarrow state

θ : learning variables \leftrightarrow control

u : given input data

~ simplify notation:

$$\min_{y, \theta} J(y, \theta) \quad \text{such that} \quad y^{[L]} = \mathcal{F}(\theta)$$

Lagrange functional:

$$\mathcal{L}(y, \theta; \phi) = J(y, \theta) - \underbrace{\langle y^{[L]} - \mathcal{F}(\theta), \phi \rangle}_{= \sum_{i=1}^L \langle y^{[i]} - f_{i-1}(y^{[i]}, \theta^{(i-1)}), \phi^{[i]} \rangle}$$

ϕ : adjoint variable

KKT conditions

- (KKT 1) $\mathcal{L}_y(\bar{y}, \bar{\theta}; \bar{\Phi})y = 0$

~ backward propagation \leftrightarrow (AE)

$$\Phi^{[l]} = \partial_{y^{[l]}} f_l(y^{[l]}, \theta^{[l]}) , l=1, \dots, L$$

$$\Phi^{[l]} = \partial_{y^{[l]}} J(y, \theta)$$

- (KKT 2) $\mathcal{L}_{\theta}(\bar{y}, \bar{\theta}; \bar{\Phi})\theta = 0$

- (KKT 3) $\mathcal{L}_{\Phi}(\bar{y}, \bar{\theta}; \bar{\Phi})\Phi = 0$

~ forward propagation \leftrightarrow (SE)

$$y^{[l]} = f_{l-1}(y^{[l-1]}, \theta^{[l-1]}) , l=1, \dots, L$$