Shallow Ritz for 1D Diffusion Equation (C.-Doktorova-Falgout-Herrera)

problem

$$\begin{cases} -(au')' = f & \text{in } I=(0,1) \\ u(0)=0 \text{ and } & a(1)u'(1)=\beta \end{cases}$$

$$\int_0^1 fv\,dx = -\int_0^1 (au')'v\,dx = \int_0^1 au'v'\,dx - au'v\Big|_0^1 = \int_0^1 au'v'\,dx - \beta v(1)$$

assuming $v(0)=0$

Let $a(u,v)=\int_0^1 au'v'\,dx$ and $f(v)=\int_0^1 fv\,dx+\beta v(1)$

$$V_0=\{v\in H^1(I) \mid v(0)=0\} \text{ and } V=H^1(I)$$

weak formulation Find $u\in V\cap\{u(0)=0\}=V_0$ s.t.

$$a(u,v)=f(v) \quad \forall v\in V_0.$$

Ritz formulation Find $u\in V_0$ s.t.

$$J(u)=\min_{v\in V_0} J(v), \text{ where } J(v)=\frac{1}{2}a(v,v)-f(v)$$

Shallow ReLU NN

$$M_n(I)=\left\{ \sum_{i=0}^{n} c_i\sigma(x-b_i) \;\middle|\; c_i\in\mathbb{R},\ 0=b_0<b_1<\cdots<b_n<1 \right\}$$

Shallow Ritz method Find $u_n(x)=\sum_{i=0}^{n} c_i\sigma(x-b_i)\in M_n(I)$ s.t.

$$J(u_n)=\min_{v\in M_n(I)} J(v).$$

Error Estimate

- the energy norm $\|v\|_a = \int_0^1 a\left(v'(x)\right)^2 dx$

- the Poincaré Inequality $\forall\, v \in V_0 = \left\{ v \in H^1(I) \,\middle|\, v(0)=0 \right\}$

  $\exists$ a constant $C>0$ s.t. $\|v\|_{0,I}^2 \equiv \int_0^1 v^2(x)\,dx \le C \|v\|_a^2$.

Proof $v(0)=0 \Rightarrow v(x) = \int_0^x v'(\xi)\,d\xi$

$$|v(x)|^2 = \left| \int_0^x v'(\xi)\,d\xi \right|^2 = \left| \int_0^x 1 \cdot v'(\xi)\,d\xi \right|^2$$

$$\le \left(\int_0^x 1^2\,d\xi \right) \left(\int_0^x v'(\xi)^2\,d\xi \right)$$

the Cauchy-Schwarz ineq.

$$\int_0^1 fg \le \left(\int_0^1 f^2 \right)^{\frac{1}{2}} \left(\int_0^1 g^2 \right)^{\frac{1}{2}}$$

$$\le x \int_0^1 v'(\xi)^2\,d\xi$$

$$\le \frac{x}{a_0} \int_0^1 a\, v'(\xi)^2\,d\xi \qquad a(x) \ge a_0 > 0 \quad \forall x \in I$$

$$\Rightarrow a_0 \|v\|_{0,I}^2 \le \int_0^1 x\,dx\, \|v\|_a^2 = \frac{1}{2} \|v\|_a^2 \Rightarrow C = \frac{1}{2a_0}.$$

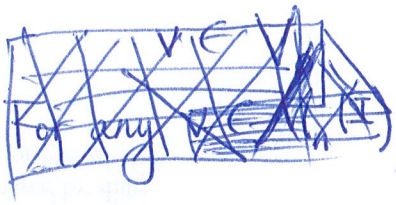
- error identity $\|u-u_n\|_a^2 = 2\left(J(u_n) - J(u)\right)$

Proof First, $a(u,u) = -2J(u)$

$$a(u-u_n, u-u_n) = a(u_n,u_n) - 2a(u,u_n) + a(u,u)$$

$$\Rightarrow \|u-u_n\|_a^2 = \left[a(u_n,u_n)-2J(u_n)\right]+2\left[J(u_n)-a(u,u_n)\right]-2J(u)$$

$$= 2\left(J(u_n)-J(u)\right)$$



Lemma Let u be the solution of WF. For any $v \in V_n$ (I),

$$\Rightarrow \|u-v\|_a^2 = 2\left(J(v)-J(u)\right), \quad \forall\, v \in V_0.$$

Lemma Let u and u_n be the solutions of WF and the shallow Ritz method. Then

$$\|u-u_n\|_a \leq \min_{v\in V_0} \|u-v\|_a \leq \min_{\vec{b}} \|u-u_I\|_a \lesssim C\left(\frac{1}{n}\right).$$

where u_I is the interpolant of u on the optimal mesh $\vec{b}^{opt}$.

Proof $\|u-u_n\|_a^2 = 2\left(J(u_n)-J(u)\right)$ $\qquad J(u_n) \leq J(v) \ \forall v \in V_0$

$$\leq 2\left(J(v)-J(u)\right) = \|u-v\|_a^2, \quad \forall v \in V_0$$

$$\Rightarrow \|u-u_n\|_a^2 \leq \min_{v\in V_0} \|u-v\|_a^2.$$

Optimality Conditions

Let $H(t)$ and $\delta(t)$ be the Heaviside step and the Dirac delta functions

$$H(t)=\begin{cases}0, & t<0\\ \frac{1}{2}, & t=0\\ 1, & t>0\end{cases} \quad \text{and} \quad \delta(t)=\begin{cases}\infty, & t=0,\\ 0, & t\neq 0\end{cases} \text{ such that } \int_{-\infty}^{\infty}\delta(t)\,dt=1.$$

Remark $H(t)=D_w\,\delta(t)=\delta'(t)$ and $\delta(t)=D_w H(t)=H'(t)=\delta''(t)$.

Set

$$\vec{\sum}(x;\vec{b})=\begin{pmatrix}\sigma(x)\\ \sigma(x-b_1)\\ \vdots\\ \sigma(x-b_n)\end{pmatrix}_{(n+1)\times 1},\quad \vec{H}(x;\vec{b})=\begin{pmatrix}H(x-b_1)\\ \vdots\\ H(x-b_n)\end{pmatrix},\quad \vec{\Lambda}(x)=\begin{pmatrix}\delta(x-b_1)\\ \vdots\\ \delta(x-b_n)\end{pmatrix}$$

and $\hat{c}=(c_0,c_1,\cdots,c_n)^T$, $\vec{c}=(c_1,\cdots,c_n)$, $\vec{b}=(b_1,\cdots,b_n)$

with these notations, let

$$u_n(x)=\sum_{i=0}^{n}c_i\,\sigma(x-b_i)=\vec{\sum}^{T}(x;\vec{b})\,\hat{c}.$$

Then we have

$$u_n(1)=\sum_{i=0}^{n}c_i\,\sigma(1-b_i)=\sum_{i=0}^{n}c_i(1-b_i)=\vec{\sum}^{T}(1;\vec{b})\,\hat{c}$$

$$u_n'(x)=\sum_{i=0}^{n}c_i\,H(x-b_i)=c_0+\vec{H}^{T}(x;\vec{b})\,\vec{c}$$

$$u_n'(b_j)=\sum_{i=0}^{n}c_i\,H(b_j-b_i)=\sum_{i=0}^{j-1}c_i+\frac{1}{2}c_j$$

$$\nabla_{\hat{c}} u_n(x) = \nabla_{\hat{c}} \vec{\Sigma}^T(x;\vec{b})\,\vec{c} = \vec{\Sigma}(x;\vec{b})$$

$$\nabla_{\hat{c}} u_n'(x) = \begin{pmatrix} 1 \\ \vec{H}(x;\vec{b}) \end{pmatrix}, \quad \nabla_{\hat{c}} u_n(1) = \vec{\Sigma}(1;\vec{b}) = \begin{pmatrix} 1 \\ 1-b_1 \\ \vdots \\ 1-b_n \end{pmatrix}$$

$$\nabla_{\vec{b}} u_n(x) = -\left(c_1 H(x-b_1), \cdots, c_n H(x-b_n)\right)^T = -D(\vec{c})\,\vec{H}(x;\vec{b})$$

where $D(\vec{c}) = \mathrm{diag}(c_1, \cdots, c_n)$ is a diagonal matrix.

$$\nabla_{\vec{b}} u_n(1) = -D(\vec{c})\,\vec{H}(1;\vec{b}) = -D(\vec{c})\,\vec{1}, \quad \text{where } \vec{1} = (1, \cdots, 1)^T.$$

$$\nabla_{\vec{b}} u_n'(x) = -\left(c_1 \delta(x-b_1), \cdots, c_n \delta(x-b_n)\right)^T = -D(\vec{c})\,\vec{\Lambda}(x;\vec{b})$$

Lemma (1) $\nabla_{\hat{c}} J(u_n) = A(\vec{b})\,\vec{c} - \vec{F}(\vec{b})$

where $A(\vec{b}) = \int_0^1 a \begin{pmatrix} 1 \\ \vec{H} \end{pmatrix} (1, \vec{H}^T)\, dx$ and $\vec{F}(\vec{b}) = \int_0^1 f\,\vec{\Sigma}(x;\vec{b})\,dx + \beta\left[\vec{1} - \begin{pmatrix} 0 \\ \vec{b} \end{pmatrix}\right]$

(2) $\nabla_{\vec{b}} J(u_n) = D(\vec{c})\,\vec{g}(\vec{b})$

where $\vec{g}(\vec{b}) = -\int_0^1 a\,u_n'(x)\,\vec{\Lambda}(x;\vec{b})\,dx + \int_0^1 f\,\vec{H}(x;\vec{b}) + \beta\vec{1}$

$$= \begin{pmatrix} \int_{b_1}^1 f(x)dx + \beta - a(b_1)\,u_n'(b_1) \\ \vdots \\ \int_{b_n}^1 f(x)dx + \beta - a(b_n)\,u_n'(b_n) \end{pmatrix}$$

Let $g_i(b_i) = \int_{b_i}^1 f\,dx + \beta - a(b_i)\,u_n'(b_i)$

Proof

(1) $$\nabla_{\hat{c}} J(u_n) = \nabla_{\hat{c}} \left[\frac{1}{2} \int_0^1 a (u_n')^2 dx - \int_0^1 f u_n dx - \beta u_n(1) \right]$$

$$= \int_0^1 a u_n' \nabla_{\hat{c}} u_n'(x) dx - \int_0^1 f \nabla_{\hat{c}} u_n(x) dx - \beta \nabla_{\hat{c}} u_n(1)$$

$$= \underbrace{\int_0^1 a \left[\vec{H}(x;\vec{b}) \right]^T \left(1, \vec{H}(x;b) \right) \vec{c}}_{A(\vec{b})} - \underbrace{\left[\int_0^1 f \vec{\Sigma}(x;\vec{b}) dx + \beta \vec{\Sigma}(1;\vec{b}) \right]}_{\vec{F}(\vec{b})}$$

(2) $$\nabla_{\vec{b}} J(u_n) = \int_0^1 a u_n' \nabla_{\vec{b}} u_n' - \int_0^1 f \nabla_{\vec{b}} u_n - \beta \nabla_{\vec{b}} u_n(1)$$

$$= -\int_0^1 D(\vec{c}) \vec{\Lambda}(x;\vec{b}) a u_n'(x) dx + \cancel{\int_0^1 f D(\vec{c}) \vec{H}(x;\vec{b}) dx} \; D(\vec{c}) \left[\int_0^1 f \vec{H} + \beta \vec{1} \right]$$

$$= D(\vec{c}) \underbrace{\left[\int_0^1 f \vec{H}(x;\vec{b}) dx + \beta \vec{1} - \int_0^1 a u_n'(x) \vec{\Lambda}(x;\vec{b}) \right]}_{\vec{g}(\vec{b})}$$

where $g_i(b_i) = \int_{b_i}^1 f(x) dx + \beta - a(b_i) u_n'(b_i)$

Lemma (optimality conditions)

(1) $A(\vec{b}) \vec{c} = \vec{F}(\vec{b})$

(2) $D(\vec{c}) \vec{g}(\vec{b}) = \vec{0}$

Lemma (1) $A(\vec{b})$ is sym. pos. def. providing that $b_i \neq b_j$ if $i \neq j$.

(2) $\kappa(A(\vec{b})) = O(n\, h_{min}^{-1})$ where $h_{min} = \min_{0 \le i \le n} (b_{i+1} - b_i)$.

Proof (1) $\forall\, \vec{\xi} \in R^{n+1}$, let $v(x) = \vec{\xi}^T \vec{H}$, then

$$\vec{\xi}^T A(\vec{b}) \vec{\xi} = \vec{\xi}^T \int_0^1 a \begin{pmatrix} 1 \\ \vec{H} \end{pmatrix} (1, \vec{H}^T)\, dx\, \vec{\xi} = \int_0^1 a\, v(x)^2 dx \ge 0$$

$$\vec{\xi}^T A(\vec{b}) \vec{\xi} = 0 \iff v(x) = 0 \ \forall x \in I \iff \vec{\xi} = \vec{0}$$

$\uparrow$ $\{1, H(x-b_1), \cdots, H(x-b_n)\}$ l. independence

Let $a = 1$.

(2) $$\vec{\xi}^T A \vec{\xi} = \int_0^1 \left(\sum_{i=0}^n \xi_i H(x-b_i) \right)^2 dx$$

$$\le \int_0^1 \left(\sum_{i=0}^n \xi_i^2 \right) \left(\sum_{i=0}^n H^2(x-b_i) \right) dx = |\vec{\xi}|^2 \sum_{i=0}^n \int_0^1 H^2(x-b_i)\, dx$$

$$= |\vec{\xi}|^2 \sum_{i=0}^n \int_{b_i}^1 dx = |\vec{\xi}|^2 \sum_{i=0}^n (1-b_i) < (n+1) |\vec{\xi}|^2 .$$

the lower bound

$$\vec{\xi}^T A \vec{\xi} = \sum_{j=0}^n \int_{b_j}^{b_{j+1}} \left(\sum_{i=0}^j \xi_i H(x-b_i) \right)^2 = \sum_{j=0}^n (b_{j+1} - b_j) \left(\sum_{i=0}^j \xi_i \right)^2 \ge h_{min} \sum_{j=0}^n \left(\sum_{i=0}^j \xi_i \right)^2$$

and $$|\vec{\xi}|^2 = \sum_{j=0}^n \left[\sum_{i=0}^j \xi_i - \sum_{i=0}^{j-1} \xi_i \right]^2 \le 2 \sum_{j=0}^n \left(\sum_{i=0}^j \xi_i \right)^2 + 2 \sum_{j=0}^n \left(\sum_{i=0}^{j-1} \xi_i \right)^2$$

$$\le 4 \sum_{j=0}^n \left(\sum_{i=0}^j \xi_i \right)^2 \le \frac{4}{h_{min}} \vec{\xi}^T A \vec{\xi}$$

$$\Rightarrow \kappa(A) \le \frac{n+1}{(h_{min}/4)} = \frac{1}{4}(n+1)\, h_{min}^{-1} .$$

Lemma If $a(x)$ is differentiable at $\{b_i\}_{i=1}^n$, then the Hessian is

$$\mathcal{H}(\vec{c},\vec{b}) = \nabla_b^2 J(u_n) = D(\vec{g}')D(\vec{c})$$

where $D(\vec{g}') = \operatorname{diag}\left(g_1'(b_1), \cdots, g_n'(b_n)\right)$

and $g_i'(b_i) = -f(b_i) - a'(b_i)u_n'(b_i)$ for $i=1,\cdots,n$.

Proof $\mathcal{H}(\vec{c},\vec{b}) = \nabla_b\left(\nabla_b J(u_n)\right)^T = \nabla_b \vec{g}^T D(\vec{c})$

$$= \left(\nabla_b g_1(b_1), \cdots, \nabla_b g_n(b_n)\right) D(\vec{c}) = \begin{pmatrix} g_1'(b_1) & & \\ & \ddots & \\ & & g_n'(b_n) \end{pmatrix} D(\vec{c}).$$

Lemma If $c_i \neq 0$ and $g_i'(b_i) \neq 0$, then

$$\mathcal{H}^{-1}(\vec{c},\vec{b}) = \operatorname{diag}\left(\left[c_1 g_1'(b_1)\right]^{-1}, \cdots, \left[c_n g_n'(b_n)\right]^{-1}\right).$$

Condition: $c_i = 0$

(i) $u_n(x) = \sum_{j=0}^{i-1} c_j \sigma(x-b_j) + \sum_{j=i+1}^{n} c_j \sigma(x-b_j)$

$\Rightarrow$ the i^{th} neuron has no contribution to $u_n(x)$.

(ii) $u_n(x)$ is piecewise linear w.r.t. partition $b_0 < b_1 < \cdots < b_n < b$

$$u_n'\left(\frac{b_i+b_{i+1}}{2}\right) = \sum_{j=0}^{i} c_j \quad \text{and} \quad u_n'\left(\frac{b_{i-1}+b_i}{2}\right) = \sum_{j=0}^{i-1} c_j$$

$$\Rightarrow c_i = u_n'\left(\frac{b_i+b_{i+1}}{2}\right) - u_n'\left(\frac{b_{i-1}+b_i}{2}\right) = u_n'(b_i^+) - u_n'(b_i^-)$$

$c_i = 0 \Rightarrow u_n(x)$ is linear in $(b_{i-1}, b_{i+1}) \Rightarrow b_i$ is not needed.

Condition: $g_i'(b_i) = 0$

$$0 = g_i'(b_i) = -f(b_i) - a'(b_i)\, u_n'(b_i)$$

$$\approx -f(b_i) - a'(b_i)\, u'(b_i) = a(b_i)\, u''(b_i)$$

$$\Rightarrow u''(b_i) \approx 0$$

Definition (1) b_i is an inflection pt of $u(x)$

$\iff u''(b_i) = 0$ and $u(x)$ changes convexity passing b_i

(2) b_i is an undulation pt of $u(x)$

$\iff u''(b_i) = 0$ and $u(x)$ does not change convexity.

~~An inflection pt~~

Remark (1) If $|c_i| \le \tau_1$ and $|g_i'(b_i)| \le \tau_2$ for relative small τ_1 and τ_2, then b_i is an almost inflection pt and b_i is not needed.

(2) If $|c_i| > \tau_1$ and $|g_i'(b_i)| \le \tau_2$, then b_i is possibly an undulation pt and b_i is fixed and no update.