HWLesson3: Deep Neural Networks

Problem 1 Given a uniform mesh $\mathcal{T}_h$ of $[0,1]$ with mesh size h, define

$$V_h = \{v : v \text{ is continuous and piecewise linear w.r.t.} \mathcal{T}_h, v(0) = v(1) = 0\}.$$

Consider the following optimization problem: Find $u_h \in V_h$ such that

$$u_h = \underset{v_h \in V_h}{\arg\min}\, J(v_h). \tag{1}$$

where $f = -\pi^2 \sin \pi x$ and $J(v_h) = \frac{1}{2}\int_0^1 |v_h'|^2 dx - \int_0^1 f v_h dx$.

1. (**10 %**) Let $u_h \in V_h$ be given by (1). For any $v_h \in V_h, t \in R$, consider the following auxiliary function

   $$g(t) = J(u_h + t v_h). \tag{2}$$

   (a) Prove that $g(0) = \min_{t \in \mathbb{R}^1} g(t)$.

   (b) Verify that
   $$g'(0) = \int_0^1 u_h' v_h' dx - \int_0^1 f v_h dx. \tag{3}$$

   (c) Prove that
   $$\int_0^1 u_h' v_h' dx = \int_0^1 f v_h dx \quad v_h \in V_h. \tag{4}$$

2. (**10 %**) Use gradient descent method to solve problem (1) for $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ and plot the corresponding solution u_h.

Problem 2 (25 %) Consider a sequence of uniform grid on interval $[0,1]$ as follows:

$$0 = x_0^k < x_1^k < \cdots < x_{N_k+1}^k = 1, \quad x_j^k = \frac{j}{N_k + 1}, \quad k = 1, 2, 3,$$

with $h_k = 1/(N_k + 1)$ and $N_k = 2^k - 1$. On each level k, we have the standard linear finite element space V_h^k with standard nodal basis functions ϕ_i^k. Denote all the interior nodes x_j^k, $j = 1, 2, \cdots, N_k$ on level k by $\mathcal{N}_k$, the so-called *hierarchical basis* (HB) refers to a special set of nodal basis functions

$$\{\phi_i^k : x_i^k \in \mathcal{N}_k \backslash \mathcal{N}_{k-1},\ k = 1, 2, 3\}.$$

We shall denote the scaled HB by $\{\psi_i\}$, $i = 1, 2, \cdots, 7$ as shown in Figure 1. Show that $\int_0^1 \psi_i' \psi_j' dx = 0$ if $i \neq j$.

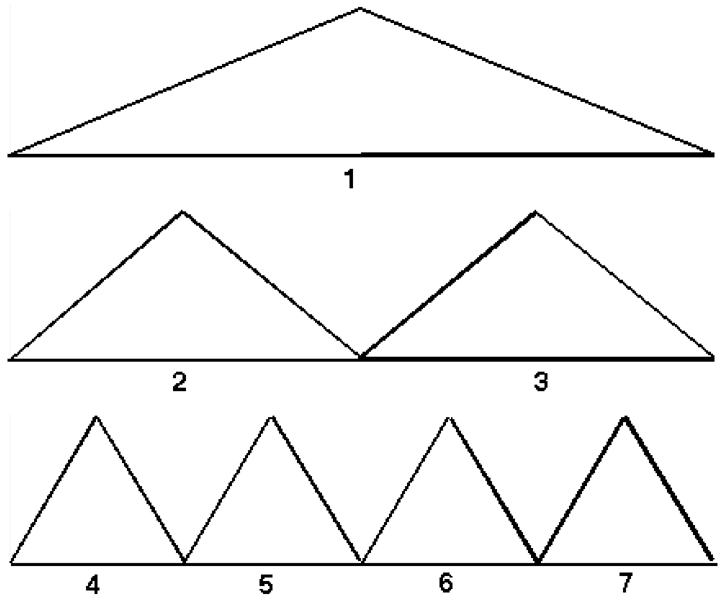


Figure 1: One dimensional Hierarchical basis

Problem 3 (**15 %**) Given a uniform mesh $\mathcal{T}_h$ of $[0, 1]$ with mesh size h, define finite element space

$$V_h = \{v : v \text{ is continuous and piecewise linear w.r.t.} \mathcal{T}_h\}.$$

Given $u(x) = x^2$, we define its finite element interpolation, $u_I \in V_h$, as follows:

$$u_I(x) = \sum_{i=1}^{N} u(x_i)\phi_i(x), \tag{5}$$

where $\{\phi_i(x)\}_{i=1}^N$ are the nodal basis functions of V_h.

Prove that

$$\|u - u_I\| \leq 16h^2.$$

Problem 4 Consider the activation function $\text{ReLU}(x) = \max(0, x)$. Answer the next questions about deep neural networks with ReLU activation function on $\mathbb{R}$.

1. **(5 %)** Consider the next deep neural network function with one hidden layer

$$g(x) = a_1\text{ReLU}(x) + a_2\text{ReLU}(x - \frac{1}{2}) + a_3\text{ReLU}(x-1) \quad \forall x \in \mathbb{R}.$$

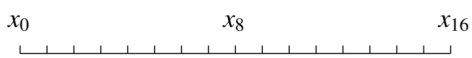
Find a_1, a_2, a_3 such that the function $g(x)$ has the following explicit formulation

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

Hint: Check this one interval by one interval.

2. **(5 %)** Given a uniform partition $\mathcal{T}_{16}$ on $[0, 1]$:

$$0 = x_0 < x_1 < \cdots < x_{16} = 1, \quad x_j = \frac{j}{16}\ (j = 0, 1, \cdots, 16).$$



And the finite element space

$$V_{16} = \{v : v \text{ is continuous and piecewise linear w.r.t.} \mathcal{T}_{16}\}.$$

Consider a $v(x) \in V_{16}$ where $v(x_j) = a_j$ for some $a_j \in \mathbb{R}$ and $j = 0, 1, \cdot, 16$, please find c_j for $j = 0, 1, \cdot, 16$ such that

$$v(x) = \sum_{j=0}^{16} c_j \text{ReLU}(16(x - x_j)), \quad \forall x \in [0, 1].$$

Hint: Check the identity by intervals from $[x_0, x_1]$ to $[x_{15}, x_{16}]$.

3. **(5 %)** For a more compact formula of $g(x)$ as a one hidden neural network, please find $V = (v_1, v_2, v_3) \in \mathbb{R}^{1\times 3}$, $W = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^{3\times 1}$ and $c \in \mathbb{R}$ such that

$$g(x) = V\text{ReLU}(Wx + b) + c.$$

Hint: Write the definition of $g(x)$ in a matrix version.

4. **(5 %)** Plot functions g_2 and $g_3(x)$ for $x \in [0, 1]$. Here $g_2(x)$ and $g_3(x)$ are defined as

$$g_2(x) = g \circ g(x) = g(g(x)), \quad \forall x \in [0, 1],$$
$$g_3(x) = g \circ g \circ g(x) = g(g(g(x))), \quad \forall x \in [0, 1].$$

Hint: Check this by intervals.

5. **(5 %)** Prove that this function $g_2(x)$ can be represented by a deep neural network with two hidden layers. In other words, please find (W_i, b_i) for $i = 1, 2, 3$ such that

$$g_2(x) = W_3\text{ReLU}\,(W_2\text{ReLU}\,(W_1 x + b_1) + b_2) + b_3, \quad \forall x \in [0, 1].$$

Hint: Recall that $g(x) = V\text{ReLU}(Wx + b) + c$ and $g_2(x) = g \circ g(x)$.

Problem 5 Consider data sets $A_1, A_2 \subset \mathbb{R}^2$ defined by:

$$A_1 = \left\{(0, \sqrt{2}), (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), (\sqrt{2}, 0), (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), (-\sqrt{2}, 0), (-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}), (0, -\sqrt{2}), (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})\right\},$$

and

$$A_2 = \left\{(0, \sqrt{5}), (\sqrt{2}, \sqrt{2}), (\sqrt{5}, 0), (-\sqrt{2}, \sqrt{2}), (-\sqrt{5}, 0), (-\sqrt{2}, -\sqrt{2}), (0, -\sqrt{5}), (\sqrt{2}, -\sqrt{2})\right\}.$$

1. **(5 %)** Consider $\varphi(x_1, x_2) = 3 - x_1^2 - x_2^2$, please verify that

   $$\varphi(x_1, x_2) > \frac{1}{2} \quad \forall (x_1, x_2) \in A_1 \quad \text{and} \quad \varphi(x_1, x_2) < -\frac{1}{2} \quad \forall (x_1, x_2) \in A_2.$$

2. **(10 %)** Prove that there exists a one hidden layer ReLU DNN function

   $$\tilde{\varphi}(x_1, x_2) = \sum_{i=1}^{n} a_i \sigma(w_i \cdot (x_1, x_2) + b_i),$$

   where $\sigma(x) = \text{ReLU}(x) = \max\{0, x\}$, such that

   $$\tilde{\varphi}(x_1, x_2) > 0 \quad \forall (x_1, x_2) \in A_1 \quad \text{and} \quad \tilde{\varphi}(x_1, x_2) < 0 \quad \forall (x_1, x_2) \in A_2.$$

   (That is to say A_1 and A_2 can be (nonlinearly) classified via one hidden layer ReLU DNN.)

   Hint: Use the universal approximation property for one hidden layer DNN with ReLU activation function (non-polynomial) to approximate the function $\varphi(x_1, x_2)$ defined in last question.

Optional Problems

Problem Consider $\phi(x, y)$ defined by

$$\phi_2(x, y) = 2(g_2(\frac{x}{2}) + g_2(\frac{y}{2}) - g_2(\frac{x+y}{2})) \quad 0 \le x \le 1, \ 0 \le y \le 1,$$

where $g_2(x) = g \circ g(x) = g(g(x))$ is defined in the previous problems.

1. Plot the function $\phi_2(x, y)$ for $0 \le x \le 1$ and $0 \le y \le 1$.
2. Furthermore, consider

   $$\phi_3(x, y) = 2(g_3(\frac{x}{2}) + g_3(\frac{y}{2}) - g_3(\frac{x+y}{2})) \quad 0 \le x \le 1, \ 0 \le y \le 1,$$

   where $g_3(x) = g \circ g_2(x) = g \circ g \circ g(x) = g(g(g(x)))$ is defined in the previous problems. Please plot or describe the properties of $\phi_3(x, y)$.