Assignment 3

November 25, 2025

1 Introduction to Sobolev Space

Background

In the study of Partial Differential Equations (PDEs), the classical function spaces C^k (continuously differentiable functions) and L^p (Lebesgue spaces) often prove insufficient. The **Sobolev space** emerged as a crucial tool, providing a framework for analyzing PDEs by incorporating the concept of **weak derivatives**.

Definition of Sobolev Space

The Sobolev space $W^{k,p}(\Omega)$ is defined as the set of functions $u \in L^p(\Omega)$ whose weak derivatives up to order k are also in $L^p(\Omega)$:

$$W^{k,p}(\Omega) = \{ \, u \in L^p(\Omega) \; : \; D^\alpha u \in L^p(\Omega), \; |\alpha| \leq k \, \} \, .$$

Symbol Meaning

- Ω: The domain of the function u.
- u: A function defined on Ω.
- $L^p(\Omega)$: The L^p Space, where $u \in L^p(\Omega)$ if and only if $\int_\Omega |u(x)|^p \, dx < \infty$.
- $D^\alpha u$: The **weak derivative** of the function u.

Weak Derivative ($v = u'$)

Let $u \in L^1_{\text{loc}}(\Omega)$. If there exists a function $v \in L^1_{\text{loc}}(\Omega)$ such that for all test functions $\phi \in C_c^\infty(\Omega)$ (smooth functions with compact support):

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

then v is called the weak derivative of u, denoted as $v = u'$.

Multi-index Notation ($|\alpha| \leq k$)

The multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ consists of non-negative integers $\alpha_i \geq 0$. The order of the multi-index is $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$. The partial differential operator D^α is defined as:

$$D^\alpha u = \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}.$$

Sobolev Seminorm and Norm

Sobolev Seminorm

Measures the size of the highest-order derivatives:

$$|u|_{W^{m,p}(\Omega)} = \left(\sum_{|\alpha|=m} \|D^\alpha u\|_{L^p(\Omega)}^p \right)^{1/p}, \quad 1 \le p < \infty.$$
$$|u|_{W^{m,\infty}(\Omega)} = \max_{|\alpha|=m} \|D^\alpha u\|_{L^\infty(\Omega)}.$$

Sobolev Norm

Measures the size of the function and all its weak derivatives up to order m:

$$\|u\|_{W^{m,p}(\Omega)} = \left(\sum_{|\alpha|\le m} \|D^\alpha u\|_{L^p(\Omega)}^p \right)^{1/p}, \quad 1 \le p < \infty.$$
$$\|u\|_{W^{m,\infty}(\Omega)} = \max_{|\alpha|\le m} \|D^\alpha u\|_{L^\infty(\Omega)}.$$

2 Introduction to Paper (Approximation)

The following lemmas discuss the approximation of polynomials using shallow neural networks with the tanh activation function, where the approximation error is measured using the Sobolev norm $\|\cdot\|_{W^{k,\infty}}$.

Lemma [3.1]: Approximating Odd-Degree Monomials

Statement: Let $k \in \mathbb{N}_0$ and $s \in 2\mathbb{N} - 1$. Then it holds that for all $\epsilon > 0$ there exists a shallow tanh neural network $\Psi_{s,\epsilon} : [-M, M] \to \mathbb{R}^{\frac{s+1}{2}}$ of width $\frac{s+1}{2}$ such that

$$\max_{\substack{p\le s,\\ p \text{ odd}}} \left\| f_p - \left(\Psi_{s,\epsilon}\right)_{\frac{p+1}{2}} \right\|_{W^{k,\infty}} \le \epsilon,$$

where $f_p(x) = x^p$ is the monomial of degree p. Moreover, the weights of $\Psi_{s,\epsilon}$ scale as:

$$O\Big(\epsilon^{-s/2}\big(2(s+2)\sqrt{2M}\big)^{s(s+3)}\Big) \quad \text{for small } \epsilon \text{ and large } s.$$

Ideas and Interpretation

- **Goal**: To show that a shallow (1 hidden layer) tanh network of width $\frac{s+1}{2}$ can uniformly approximate all odd-degree monomials x^p (where $p \le s$) on the domain $[-M, M]$.
- **Measure**: The approximation error is controlled by the $\|\cdot\|_{W^{k,\infty}}$ norm, ensuring that the function and its weak derivatives up to order k are approximated well.
- **Key Insight (Remark)**: Since the Taylor expansion of $\tanh(x)$ consists only of odd-degree monomials:
$$\tanh(x) = x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315} + O(x^9)$$
it suggests that tanh is naturally suited for constructing approximations of odd functions/monomials.

Lemma [3.2]: Approximating All Monomials

Statement: Let $k \in \mathbb{N}_0$, $s \in 2\mathbb{N} - 1$, and $M > 0$. For every $\epsilon > 0$, there exists a shallow tanh neural network $\psi_{s,\epsilon} : [-M, M] \to \mathbb{R}^s$ of width $\frac{3(s+1)}{2}$ such that

$$\max_{p\le s} \|f_p - (\psi_{s,\epsilon})_p\|_{W^{k,\infty}} \le \epsilon.$$

Furthermore, the weights scale as:

$$O\Big(\epsilon^{-s/2}(\sqrt{M(s+2)})^{\frac{3s(s+3)}{2}}\Big) \quad \text{for small } \epsilon \text{ and large } s.$$

Ideas and Interpretation

- **Goal**: To show that a shallow tanh network can approximate **all** monomials x^p (both odd and even, where $p \leq s$) on $[-M, M]$.
- **Width Increase**: The required width increases to $\frac{3(s+1)}{2}$ compared to Lemma 3.1, primarily to accommodate the approximation of even-degree monomials.
- **Proof Concept**: The proof leverages the result of Lemma 3.1 by expressing even-degree monomials as a linear combination of odd-degree monomials (as suggested in the Remark about y^{2n}) and then using induction and error control.

3 Applying Lemma [3.2] to an Example

We choose the polynomial $f(x) = x^5 + x^4 + x^3 + x^2 + x + 1$ on the domain $x \in [-1, 1]$.

1. **Choose** s: Since the highest degree of $f(x)$ is 5, we choose the odd integer $s = 5$.
2. **Apply Lemma 3.2**: Lemma 3.2 guarantees the existence of a shallow tanh network $\psi_{5,\epsilon} : [-1, 1] \to \mathbb{R}^5$ with a fixed **width** of $\frac{3(5+1)}{2} = 9$. This network can simultaneously approximate all monomials x^p for $p = 1, 2, 3, 4, 5$.
3. **Construct the Approximation**: The target function is $f(x) = 1 + \sum_{p=1}^{5} x^p$. We construct the network approximation $\tilde{f}(x)$ by summing the network outputs for each monomial, plus the constant term:
$$\tilde{f}(x) = 1 + \sum_{p=1}^{5} (\psi_{5,\epsilon})_p(x).$$
4. **Error Control**: If we choose the precision δ such that $\max_{p \leq 5} \|x^p - (\psi_{5,\epsilon})_p\|_{W^{k,\infty}} \leq \delta$, the total error is controlled by the linearity of the norm:
$$\left\|f - \tilde{f}\right\|_{W^{k,\infty}} \leq \sum_{p=1}^{5} \|x^p - (\psi_{5,\epsilon})_p\|_{W^{k,\infty}} \leq 5\delta.$$
By taking $\delta = \varepsilon/5$ (or $\varepsilon/6$ if the constant term is also approximated), we ensure the total error is less than ε.
5. **Network Scale**: The network has a small, fixed width (9). The magnitude of the weights will grow rapidly, scaling roughly as $O(\varepsilon^{-5/2})$.

Conclusion: We can construct a width-9 shallow tanh network that approximates $f(x) = x^5 + x^4 + x^3 + x^2 + x + 1$ on $[-1, 1]$ with high accuracy in the Sobolev norm.

4 Unanswered Questions

Q2: Can we use Taylor expansion to approximate other functions (e.g., $\sin(x), \cos(x)$) using the same neural network approximation technique?