NORTHWESTERN UNIVERSITY

Operator Learning Theory: k-means Clustering and Fourier Neural Operators

Owen Huang
Princeton University, Department of Mathematics

Contributors: Dr. Wing Kam Liu, Sourav Saha, Jiachen Guo

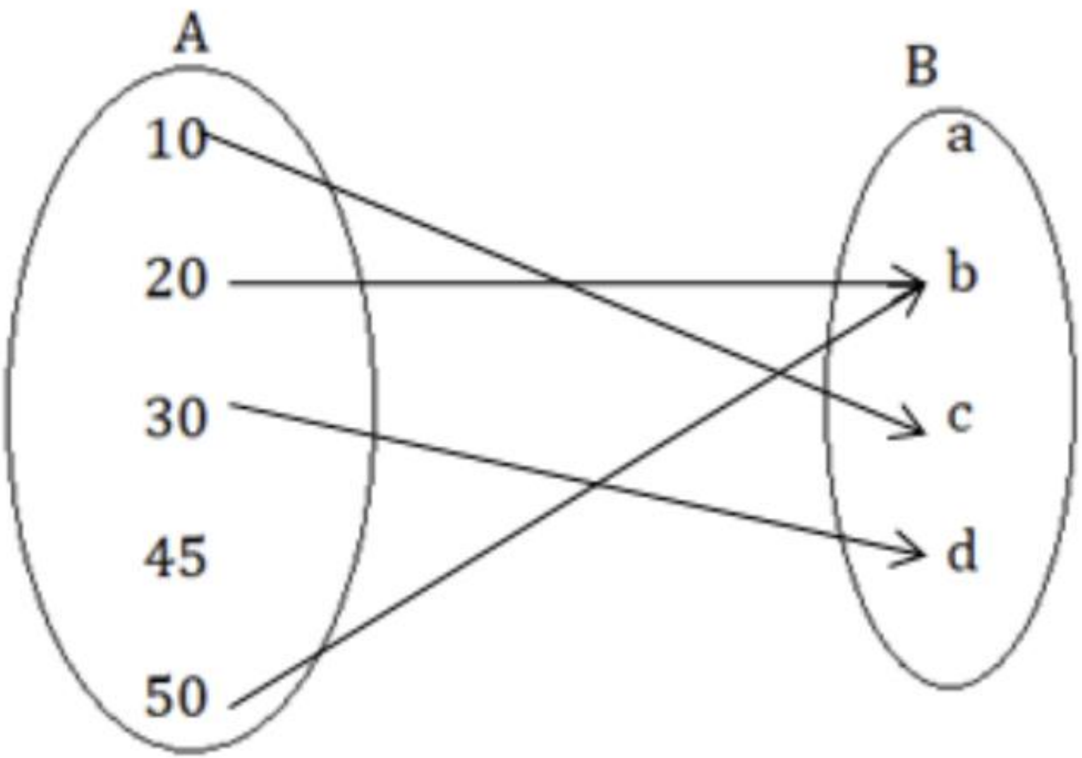
NORTHWESTERN UNIVERSITY

Outline

1. A Meta View of Neural Networks
2. Weaknesses of NN's
3. Operator Learning and Neural Operators
4. Fourier Neural Operators
5. A Mathematical Formulation of k-means Clustering
6. Conjecture and Proposed Solution
7. Future Research

A Meta View of Neural Networks

It is often helpful to speak about functions as *maps* between two collections of objects. That is, a function f takes an object from a set A and maps it to an object from set B. We write this as $f : A \to B$.



A Meta View of Neural Networks

We can thus formulate the objective of neural networks as follows. Given a secret function $f : A \to B$, train a network to find function $f^\star : A \to B$ so that

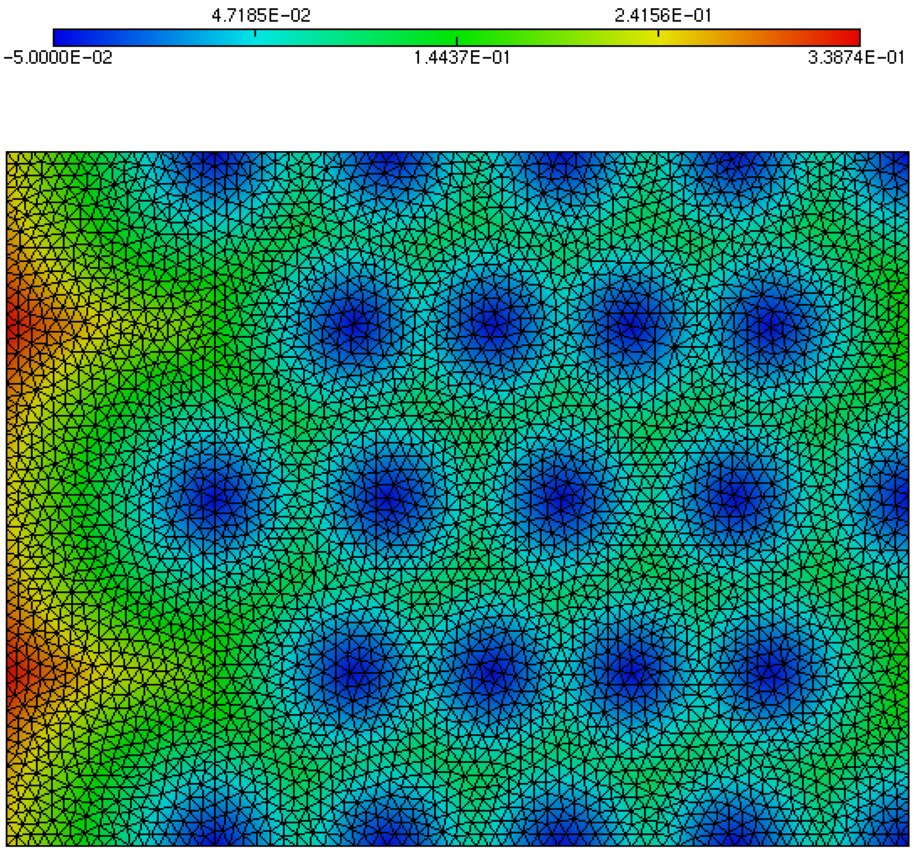
$$f(x) \approx f^\star(x)$$

for every element $x \in A$. Examples include:

1. Classification: f : {input pixels and colours} $\to$ {cat, dog, car, ...}
2. AI-driven ads: f : {previous purchases} $\to$ {possible future purchases}
3. Self-driving cars: f : {visual data} $\to$ {possible car maneuvers}

Weaknesses of Neural Networks

A relevant weakness in using NN's is its sensitivity to the way data is inputted. For example, say you are trying to examine the microstructure of a sample of a material. One may first decompose the sample into a 100×100 grid and run a CNN.



NORTHWESTERN UNIVERSITY

Weaknesses of Neural Networks

However, we would like the freedom to change the grid size to capture different features of the material (think: more resolution for finer details, and less resolutions for coarser features.) Most NN architectures do not allow for this kind of freedom, and data needs to be trained again from scratch. NN's do not need to be more computationally expensive than they already are.

Layer Type	Complexity per Layer	Sequential Operations	Maximum Path Length
Self-Attention	$O(n^2 \cdot d)$	$O(1)$	$O(1)$
Recurrent	$O(n \cdot d^2)$	$O(n)$	$O(n)$
Convolutional	$O(k \cdot n \cdot d^2)$	$O(1)$	$O(log_k(n))$
Self-Attention (restricted)	$O(r \cdot n \cdot d)$	$O(1)$	$O(n/r)$

Operator Learning and Neural Operators

A recently studied remedy is to learn the underlying *operator* instead of an just instance of it (which is what is really happening when a discretization is fixed).

An **operator** is a function $\Phi : A \to B$ where A and B are themselves sets of functions. For example,

1. In computer science, "higher level functions"
2. Differentiation
3. Convolution

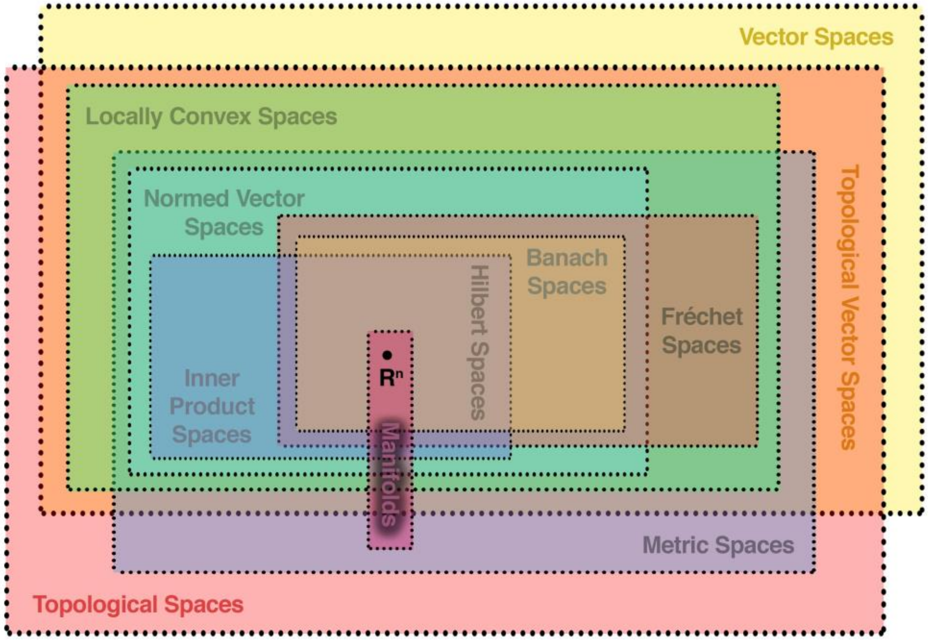
$$\frac{df}{dx} : \{\text{differentiable functions}\} \to \{\text{functions}\}$$

$$f(x) \mapsto f'(x)$$

However, we may want to impose restrictions on what functions we allow in the sets A and B.

For example, in differentiation, we can only differentiate *differentiable* functions. Other restrictions may include:

1. continuous
2. integrable
3. bounded
4. polynomial
5. linear



Operator Learning and Neural Operators**

In my research, a useful restriction is to **Sobolev functions**. Intuitively, they are a generalization of differentiable functions.

A little more formally, they are a class of functions whose p-th power is integrable and has "nice" partial derivatives.

8.2 The Sobolev Space $W^{1,p}(I)$

Let $I = (a, b)$ be an open interval, possibly unbounded, and let $p \in \mathbb{R}$ with $1 \leq p \leq \infty$.

Definition. The *Sobolev space* $W^{1,p}(I)$[1] is defined to be

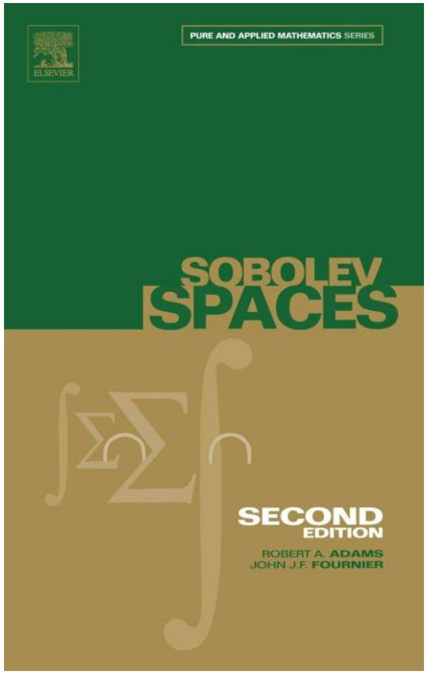
$$W^{1,p}(I) = \left\{ u \in L^p(I);\ \exists g \in L^p(I) \text{ such that } \int_I u\varphi' = -\int_I g\varphi \quad \forall \varphi \in C_c^1(I) \right\}.$$

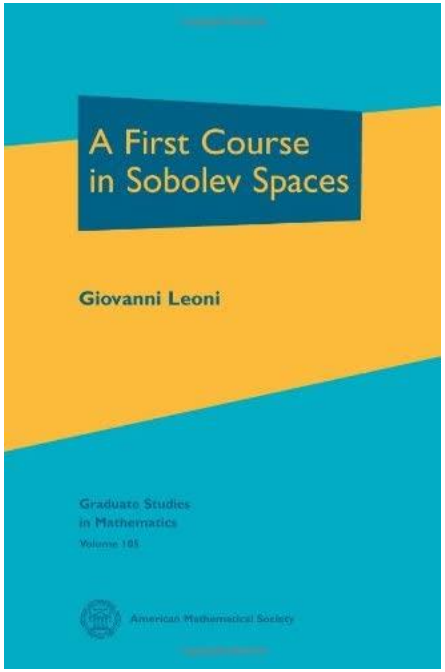
We set

$$H^1(I) = W^{1,2}(I).$$

For $u \in W^{1,p}(I)$ we denote [2] $u' = g$.

Remark 1. In the definition of $W^{1,p}$ we call φ a *test function*. We could equally well have used $C_c^\infty(I)$ as the class of test functions because if $\varphi \in C_c^1(I)$, then $\rho_n \star \varphi \in C_c^\infty(I)$ for n large enough and $\rho_n \star \varphi \to \varphi$ in C^1 (see Section 4.4; of course, φ is extended to be 0 outside I).

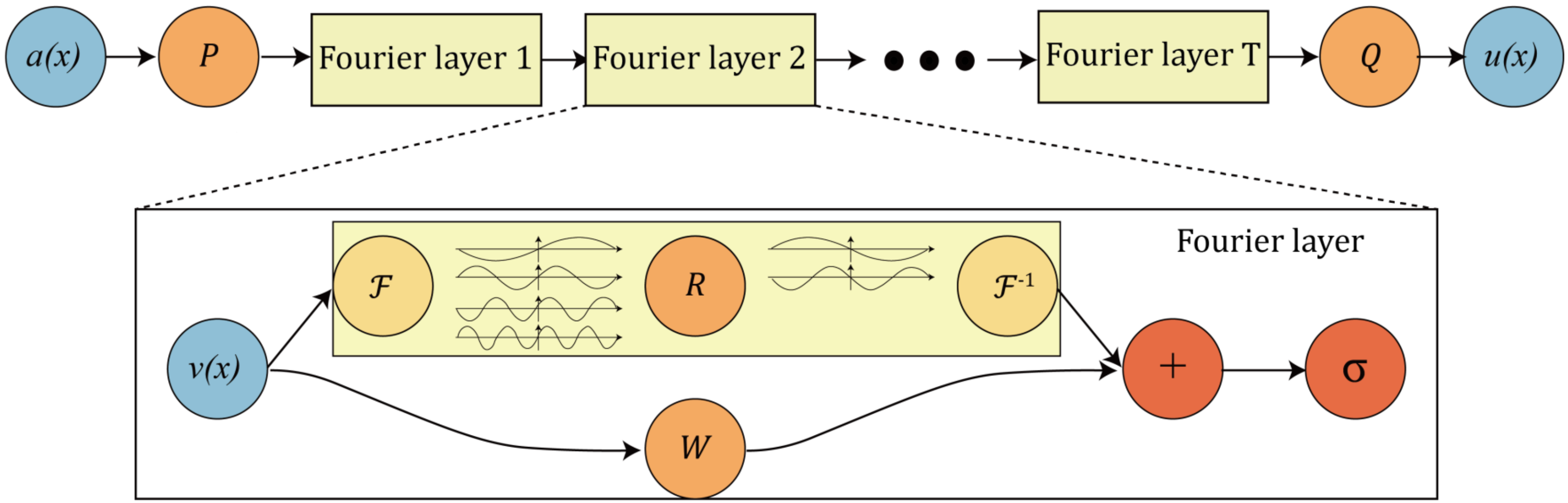




Fourier Neural Operators

A particular architecture of neural operators are called **Fourier Neural Operators** (think: convNets, RNN's, FFNN's, etc.) They are inspired by the study of differential equations and Fourier analysis.

The distinguishing components of FNO's are called *Fourier layers*, and are just learnable kernel functions - just like kernels in a convNet.



Fourier Neural Operators

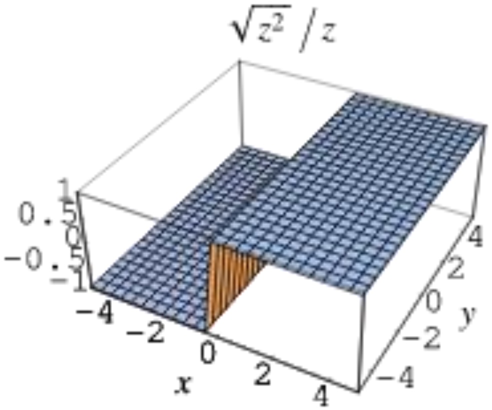
However, in some sense, FNO's have only been proven to work when learning operators between Sobolev spaces, not all functions! So if we are to use FNO's, we best make sure the operator we are learning is a Sobolev function...

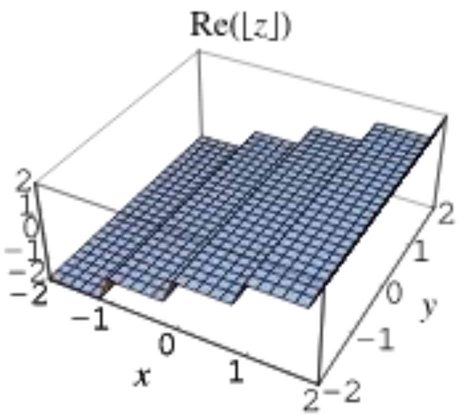
Theorem 1 (Kovachki et al., 2017). *Let $s \geq 0$, and $\Phi : H^s(D;\mathbb{R}^n) \to H^s(D;\mathbb{R}^k)$ be a continuous operator, and $\Omega \subset H^s(D;\mathbb{R}^n)$ be compact. For every $\epsilon > 0$, there exists a Fourier neural network $\mathcal{N} : H^s(D;\mathbb{R}^n) \to H^s(D;\mathbb{R}^k)$ (which can be regarded as a continuous operator) such that*

$$\sup_{a\in\Omega} \|\Phi a - \mathcal{N} a\|_{H_s} < \epsilon$$

A Mathematical Formulation of k-means Clustering

Recall the k-means clustering algorithm from class. The goal is to partition some domain Ω into k regions: $\Omega_1, \Omega_2, \ldots, \Omega_k$ based on some objective function f. After the algorithm, we can assume that f is constant within each Ω_j. This has vast uses in engineering and material science.





Self-consistent clustering analysis: an efficient multi-scale scheme for inelastic heterogeneous materials

Zeliang Liu[a], M.A. Bessa[b], Wing Kam Liu[b,*]

[a] Theoretical and Applied Mechanics, Northwestern University, Evanston, IL 60208, USA
[b] Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA

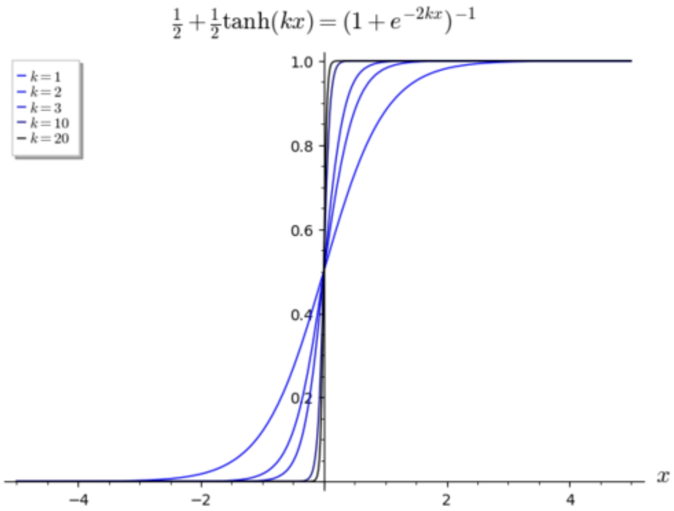
Conjecture and Proposed Solution

However, it turns out that piecewise constant functions are generally NOT Sobolev functions! Hence we cannot immediately run FNO's and expect convergence.

However, we can approximate the "clustered" function analytically (ie. exhibiting a Taylor expansion everywhere) using $\tanh(\cdot)$. This follows because if $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x \geq 0$, then

$$\lim_{k \to \infty} \frac{1}{2}(1 + \tanh(kx)) = H(x)$$

for every x.



Conjecture and Proposed Solution

As per my research, it is still an open problem if this approach works in 1 dimension, but it is our goal to prove this. We just need to consider certain nuances such as: uniform convergence, rate of convergence and ultimately verify findings using simulations.

Future Research

It will be most important to generalize this idea to multiple dimensions. Could the same approach work?

Clustering in 1 dimension always look like intervals, which are "well-behaved". However, clusters may be topologically different in higher dimensions!

Can we impose restrictions on clustering algorithms to make clusters "interval-like"?

NORTHWESTERN UNIVERSITY

Thank you all for a great summer and good luck in the future!

Questions too!

References

1) Matti Schneider, On the mathematical foundations of the self-consistent clustering analysis for non-linear materials at small strains, Computer Methods in Applied Mechanics and Engineering, Volume 354, 2019, Pages 783-80
2) , Neural operator: Graph kernel network for partial differential equations, 2020.
3) Liu, Z., Bessa, M. A., & Liu, W. K. (2016). Self-consistent clustering analysis: an efficient multi-scale scheme for inelastic heterogeneous materials. Computer Methods in Applied Mechanics and Engineering, 306, 319-341.
4) Nikola Kovachki and Samuel Lanthaler and Siddhartha Mishra on universal approximation and error bounds for fourier neural operators , Computer Methods in Applied Mechanics and Engineering, 2021
5) Tim De Ryck, Samuel Lanthaler, and Siddhartha Mishra. On the approximation of functions by tanh neural networks. Neural Networks, 143:732–750, 2021. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2021.08.015.