Universal Approximation for CNO

Guido Putignano

May 16, 2024

Explain briefly how the CNO is defined The Convolutional Neural Operator is an extension that comes from Raonic et al. [2024] that has the main function of removing the aliasing effect considering operators. This is achieved through a sequence of compositional mappings between functions, denoted as $G : u \mapsto P(u) = v_0 \mapsto v_1 \mapsto \cdots \mapsto v_L \mapsto Q(v_L) = u$.

At the outset, the input function u from the space $B_w(D)$ is elevated to a latent space of bandlimited functions v_0 via the lifting layer P.
Subsequently, the lifted function undergoes a series of mappings, with each layer comprising three elementary mappings: P_l, K_l, and Σ_l. Here, P_l may denote either an upsampling or downsampling operator, K_l represents the convolution operator, and Σ_l signifies the activation operator. Finally, the last output function v_L obtained through the iterative process is projected to the output space via the projection operator Q. Similar to the lifting operation, the projection operation is executed using a convolution operator.

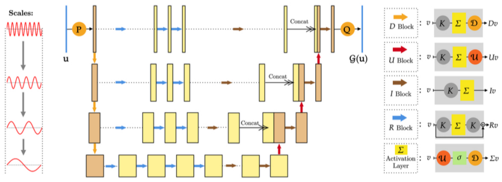

Figure 1: Convolutional Neural Operators

Explain briefly what the universality theorem means in practice. The identification and definition of The universality theorem has firstly been introduced in the paper of Kovachki et al. [2021] in an infinite dimensional setting considering FNO and later also described in the paper of Raonic et al. [2024] taking CNO into consideration. The Universality Theorem states that for any given error tolerance $\epsilon > 0$, and for any sufficiently regular solution operator $G^\dagger$, there exists a Convolutional Neural Operator (CNO) G such that the approximation error between the solution produced by the CNO and the solution produced by the true solution operator is less than ϵ for all inputs within a certain domain. Mathematically, it is expressed as:

$$\|G^\dagger(a) - G(a)\|_{L^p(D)} < \epsilon$$

where:

- a represents inputs in a suitable function space.
- $L^p(D)$ is the norm used to measure errors.
- D is the domain of the problem.

We assume that G^* is continuous. Moreover, we also assume the following modulus of continuity,

$$\|G^*(a) - G^*(a')\|_{L^p(T^2)} \leq \omega\left(\|a - a'\|_{H^\sigma(T^2)}\right),$$

for some $p \in \{2, \infty\}$ and $0 \leq \sigma \leq r - 1$, where $\omega : [0, \infty) \to [0, \infty)$ is a monotonically increasing function with $\lim_{y \to 0} \omega(y) = 0$. In this case, X^* has to be compact and G^* continuous.
Some implications are: CNOs can adapt to various PDEs and boundary conditions, CNOs can achieve precise solutions by adjusting their complexity, trained CNOs provide PDE solutions faster than traditional methods, CNOs generalize well across similar problems and CNOs handle high-dimensional and complex geometries efficiently without escalating computational complexity.

Note that we stated that universality theorem holds for sufficiently regular solution operators G†. What additional regularization constraints would you impose on G†, so that the universality theorem holds? Given the paper of Raonic et al. [2024] there are teh following condtions to impose

1. **Continuity and Lipschitz Condition:** Ensure $G^\dagger$ satisfies a Lipschitz condition to stabilize the training of CNOs:

   $$\|G^\dagger(u) - G^\dagger(v)\| \leq L\|u - v\|,$$

   where L is a Lipschitz constant.

2. **Boundedness:** $G^\dagger$ should be bounded to manage the amplitude of the output space:

   $$\|G^\dagger(u)\| \leq C\|u\| \quad \forall u \in \text{Domain of } G^\dagger.$$

3. **Smoothness:** $G^\dagger$ should map to a space of functions with sufficient smoothness:

   $$G^\dagger : H^r(D) \to H^s(D), \text{ with } s \geq r.$$

4. **Compactness:** Ensuring $G^\dagger$ is compact might be beneficial for better generalization and robustness:

   $$G^\dagger \text{ is compact} \Rightarrow \sigma(G^\dagger) \text{ has rapid decay}.$$

Make a rough sketch of the universality theorem proof The possibility of writing a proof for the universality theorem can be possible considering:

1. **Assumption of Regularity**: Assume $G^\dagger$ is continuous and satisfies a modulus of continuity, ensuring stability and smoothness in its responses. Moreover, it may be useful to consider the further constraints described above

2. **CNO Construction**: Construct G as a convolutional neural operator that maps inputs in X^* to outputs in $H^r(D)$. This involves defining the CNO to work over spaces that respect the underlying function space properties required by $G^\dagger$.

3. **Approximation in Function Spaces**: Utilize the structure of CNOs to approximate the action of $G^\dagger$ within the desired tolerance ϵ, focusing on managing and minimizing the error in the approximation of $G^\dagger$.

4. **Error Analysis**: Establish that the error between $G^\dagger(a)$ and $G(a)$ is less than ϵ by analyzing the continuity and the bounded nature of the operator $G^\dagger$ and ensuring that G captures these properties.

5. **Empirical Verification**: Use a set of benchmark problems to empirically verify that G approximates $G^\dagger$ effectively, thus reinforcing the theoretical proof with practical implementations and tests.

The mathematical demonstration of the proof of the theorem 3.1 of Raonic et al. [2024] can be found in the Appendix B of the paper.

References

Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for fourier neural operators. *Journal of Machine Learning Research*, 22(290):1–76, 2021.

Bogdan Raonic, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of pdes. *Advances in Neural Information Processing Systems*, 36, 2024.