Delft Institute of Applied Mathematics
TU Delft

Representation Equivalent Neural Operators

A framework for Alias-free Operator Learning

Francesca Bartolucci

In collaboration with

Bogdan Raonić
ETH Zürich

Emmanuel de Bézenac
INRIA

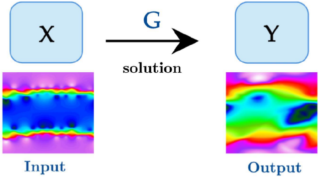
Roberto Molinaro
Jua

Rima Alaifari
RWTH Aachen

Siddhartha Mishra
ETH Zürich

Motivation

- Various scientific phenomena can be described by PDEs
- Classically, numerical methods are used to approximate solution operators
- Classical numerical methods can be computationally prohibitively expensive
- Replace classical computationally intensive algorithms with fast, robust data-driven surrogate models, which enables their use in time-critical applications



Operator learning in a nutshell

- $U\colon \mathcal{H} \to \mathcal{K}$ operator between infinite-dimensional spaces (e.g. solution operators of PDEs)
- Goal: build an approximation $U^* \approx U$ from input-output pairs

$$\{u_i, U(u_i)\}_{i=1}^N$$

- How: construct a neural operator $U_\theta\colon \mathcal{H} \to \mathcal{K}$ as a sequence of layers

$$U_\theta = U_L \circ U_{L-1} \circ \ldots \circ U_1, \qquad \theta \in \Theta,$$

and minimize

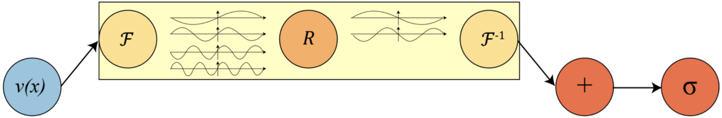
$$\frac{1}{N} \sum_{i=1}^N \|U(u_i) - U_\theta(u_i)\|_{\mathcal{K}}^2$$

- as in classical NNs, a neural operator is defined layer-wise
- ...but novelty: layers are defined as operators between infinite-dimensional spaces

Example: Fourier neural operators

- state of the art for learning solution operators of PDEs
- layer of Fourier neural operators [Li et al., ICLR 2021]:

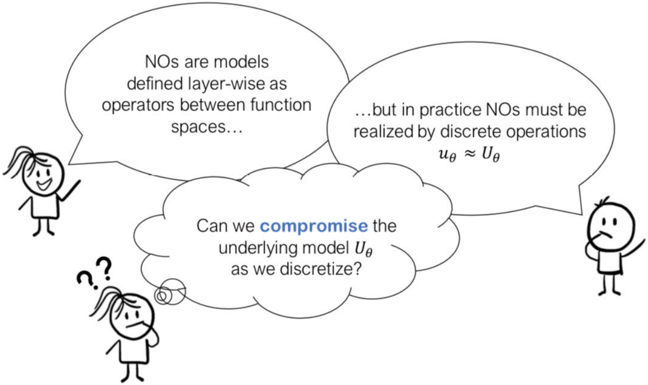
$$U_\ell = \sigma(\mathcal{F}^{-1}(\mathcal{R}_\theta \odot \mathcal{F})), \quad \ell = 1, \ldots, L,$$



- how are FNOs implemented?

The problem

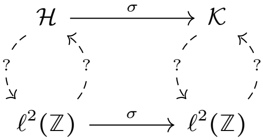
- In practice, we have only access to discrete representations of functions (e.g. grid values)...



- Loose discretizations $u_\theta \approx U_\theta$ can lead to mismatches between continuous models and their implemented discretizations, compromising the underlying model U_θ

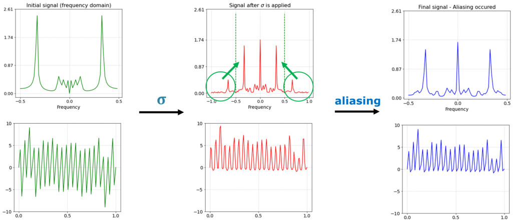
FNOs in practice: activation & aliasing

- in practice, all computations done discretely: $u_\ell = \sigma(F^{-1}(R_\theta \odot F))$
  - functions sampled on a grid: $\{f(nT)\}_{n\in\mathbb{Z}}$
  - DFT instead of Fourier transform
  - activations computed on unchanged grid: $\{\sigma(f(nT))\}_{n\in\mathbb{Z}}$



FNOs in practice: activation & aliasing

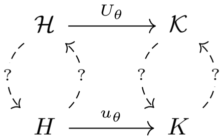
- in practice, all computations done discretely: $u_\ell = \sigma(F^{-1}(R_\theta \odot F))$
  - functions sampled on a grid: $\{f(nT)\}_{n\in\mathbb{Z}}$
  - DFT instead of Fourier transform
  - activations computed on unchanged grid: $\{\sigma(f(nT))\}_{n\in\mathbb{Z}}$



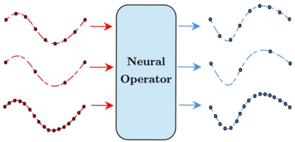
- activation functions increase bandwidth
- U_ℓ and its discretization u_ℓ are inconsistent

Consequences

- possible discrepancy between continuous and discrete levels



- instead: enforce a continuous-discrete equivalence at any resolution

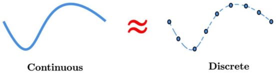


- can sampling/frame theory be leveraged to define a new class of neural operators?

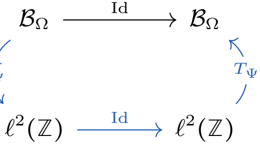
Excursion to sampling theory...

$\mathcal{B}_\Omega = \{f \in L^2(\mathbb{R}) : \text{supp}\,\hat{f} \subseteq [-\Omega, \Omega]\}$

- WSK sampling theorem: $f = \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2\Omega}\right)\ \text{sinc}(2\Omega \cdot - n)$



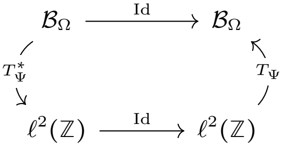
- $\Psi = \{\psi_n(x) = \text{sinc}(2\Omega x - n)\}_{n\in\mathbb{Z}}$
- synthesis operator: $T_\Psi : \ell^2(\mathbb{Z}) \to \mathcal{B}_\Omega, \quad T_\Psi(\{c_n\}_{n\in\mathbb{Z}}) = \sum_{n\in\mathbb{Z}} c_n \psi_n$
- analysis operator: $T_\Psi^* : \mathcal{B}_\Omega \to \ell^2(\mathbb{Z}), \quad T_\Psi^* f = \{\langle f, \psi_n\rangle\}_{n\in\mathbb{Z}} = \left\{f\left(\frac{n}{2\Omega}\right)\right\}_{n\in\mathbb{Z}}$
- WSK sampling theorem: $f = T_\Psi T_\Psi^* f$



Excursion to sampling theory...

- $\Psi = \{\psi_n(x) = \text{sinc}(2\Omega(x - nT))\}_{n\in\mathbb{Z}}$, $1/T < 2\Omega$
- synthesis operator: $T_\Psi : \ell^2(\mathbb{Z}) \to \mathcal{B}_\Omega, \quad T_\Psi(\{c_n\}_{n\in\mathbb{Z}}) = \sum_{n\in\mathbb{Z}} c_n \psi_n$
- analysis operator: $T_\Psi^* : \mathcal{B}_\Omega \to \ell^2(\mathbb{Z}), \quad T_\Psi^* f = \{\langle f, \psi_n\rangle\}_{n\in\mathbb{Z}} = \{f(nT))\}_{n\in\mathbb{Z}}$
- $T_\Psi T_\Psi^* f = P_{\mathcal{M}_\Psi} f, \quad \mathcal{M}_\Psi = \overline{\text{span}\{\psi_n : n \in \mathbb{Z}\}}$

The diagram



no longer commutes

- aliasing error function for sampling f at the sampling rate $1/T$

$$\varepsilon(f) = f - \mathcal{P}_{\mathcal{M}_\Psi} f$$

Excursion to sampling theory...

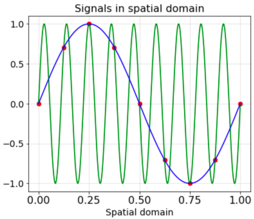
The diagram

$$
\begin{array}{ccc}
\mathcal{B}_\Omega & \xrightarrow{\mathrm{Id}} & \mathcal{B}_\Omega \\
T_\Psi^* \downarrow & & \uparrow T_\Psi \\
\ell^2(\mathbb{Z}) & \xrightarrow{\mathrm{Id}} & \ell^2(\mathbb{Z})
\end{array}
$$

no longer commutes

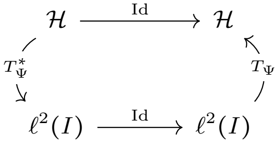
- aliasing error function for sampling f at the sampling rate $1/T$

$$\varepsilon(f) = f - \mathcal{P}_{\mathcal{M}_\Psi} f$$



Excursion to frame theory...

- $\Psi = \{\psi_i\}_{i\in I}$
- synthesis operator: $T_\Psi : \ell^2(I) \to \mathcal{H}, \quad T_\Psi(\{c_i\}_{i\in I}) = \sum_{i\in I} c_i\psi_i$
- analysis operator: $T_\Psi^* : \mathcal{H} \to \ell^2(I), \quad T_\Psi^* f = \{\langle f, \psi_i\rangle\}_{i\in I}$

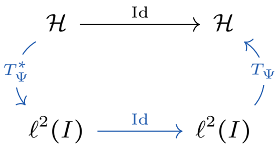


Excursion to frame theory...

- $\Psi = \{\psi_i\}_{i\in I}$ tight frame for the separable Hilbert space $\mathcal{H}$

$$\sum_{i\in I} |\langle f, \psi_i\rangle|^2 = A\|f\|^2, \qquad \forall f \in \mathcal{H}$$

- synthesis operator: $T_\Psi : \ell^2(I) \to \mathcal{H}, \quad T_\Psi(\{c_i\}_{i\in I}) = \sum_{i\in I} c_i\psi_i$
- analysis operator: $T_\Psi^* : \mathcal{H} \to \ell^2(I), \quad T_\Psi^* f = \{\langle f, \psi_i\rangle\}_{i\in I}$
- reconstruction formula: $f = \frac{1}{A} T_\Psi T_\Psi^* f = \frac{1}{A}\sum_{i\in I}\langle f, \psi_i\rangle\, \psi_i$

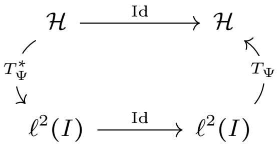


- choice of a frame = continuous-discrete equivalence between f and $\{\langle f, \psi_i\rangle\}_{i\in I}$

Excursion to frame theory...

- $\Psi = \{\psi_i\}_{i \in I}$ tight frame sequence for $\mathcal{H}$ (tight frame for $\mathcal{M}_\Psi := \overline{\text{span}_{i \in I} \{\psi_i\}}$)
- synthesis operator: $T_\Psi : \ell^2(I) \to \mathcal{H}, \quad T_\Psi(\{c_i\}_{i \in I}) = \sum_{i \in I} c_i \psi_i$
- analysis operator: $T_\Psi^* : \mathcal{H} \to \ell^2(I), \quad T_\Psi^* f = \{\langle f, \psi_i \rangle\}_{i \in I}$
- $\mathcal{P}_{\mathcal{M}_\Psi} f = \frac{1}{A} T_\Psi T_\Psi^* f = \frac{1}{A} \sum_{i \in I} \langle f, \psi_i \rangle \, \psi_i$

The diagram



no longer commutes

- aliasing error function for f w.r.t. the frame sequence $\Psi = \{\psi_i\}_{i \in I}$

$$\varepsilon(f) = f - \mathcal{P}_{\mathcal{M}_\Psi} f$$

Excursion to frame theory...

The diagram

$$\begin{array}{ccc} \mathcal{H} & \xrightarrow{\mathrm{Id}} & \mathcal{H} \\ T_{\Psi}^{*} \downarrow & & \uparrow T_{\Psi} \\ \ell^2(I) & \xrightarrow{\mathrm{Id}} & \ell^2(I) \end{array}$$

no longer commutes

- aliasing error function for f w.r.t. the frame sequence $\Psi = \{\psi_i\}_{i \in I}$

$$\varepsilon(f) = f - \mathcal{P}_{\mathcal{M}_{\Psi}} f$$

- in the presence of aliasing, continuous and discrete levels are inconsistent
- if $\varepsilon(f) \equiv 0$, we say: continuous-discrete equivalence between f and its frame coefficients
- we generalize this concepts to operators: aliasing error operator

Framework for operators

- $U\colon \operatorname{Dom} U \subseteq \mathcal{H} \to \mathcal{K}$
- $\Psi = \{\psi_i\}_{i\in I} \subseteq \mathcal{H}$ and $\Phi = \{\phi_k\}_{k\in K} \subseteq \mathcal{K}$ tight frame sequences
- $u\colon \operatorname{Ran} T_\Psi^* \to \operatorname{Ran} T_\Phi^*$

$$\begin{array}{ccc} \mathcal{H} & \xrightarrow{\quad U \quad} & \mathcal{K} \\ {\scriptstyle T_\Psi^*}\downarrow & & \uparrow{\scriptstyle T_\Phi} \\ \ell^2(I) & \xrightarrow{\quad u \quad} & \ell^2(K) \end{array}$$

- aliasing error operator of U w.r.t. the discretization u

$$\varepsilon(U,u,\Psi,\Phi) = U - T_\Phi \circ u \circ T_\Psi^*$$

- $\varepsilon(U,u,\Psi,\Phi) \equiv 0 \implies u = T_\Phi^* \circ U \circ T_\Psi$
- $\varepsilon(U,u,\Psi,\Phi) = U - \mathcal{P}_{\mathcal{M}_\Phi} \circ U \circ \mathcal{P}_{\mathcal{M}_\Psi}$
- $u = T_\Phi^* \circ U \circ T_\Psi \wedge (\operatorname{Dom} U \subseteq \mathcal{M}_\Psi \ \wedge \ \operatorname{Ran} U \subseteq \mathcal{M}_\Phi) \implies \varepsilon(U,u,\Psi,\Phi) \equiv 0$

Representation-equivalent Neural Operators

Let's go back to neural operators…build an approximation $U^* \approx U$ from input-output pairs $\{u_i, U(u_i)\}_{i=1}^N$ with a neural operator $U_\theta : \mathcal{H} \to \mathcal{K}$

$$U_\theta = U_L \circ U_{L-1} \circ \ldots \circ U_1, \qquad \theta \in \Theta.$$

For every layer $U_\ell : \mathcal{H}_\ell \to \mathcal{H}_{\ell+1}, \ell = 1, \ldots, L,$

- discretize function spaces: choose frame sequences $\Psi_\ell \subseteq \mathcal{H}_\ell,\ \Psi_{\ell+1} \subseteq \mathcal{H}_{\ell+1}$ such that

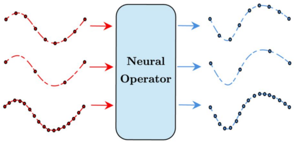
$$\operatorname{Dom} U_\ell \subseteq \mathcal{M}_{\Psi_\ell} \ \wedge\ \operatorname{Ran} U_\ell \subseteq \mathcal{M}_{\Psi_{\ell+1}}$$

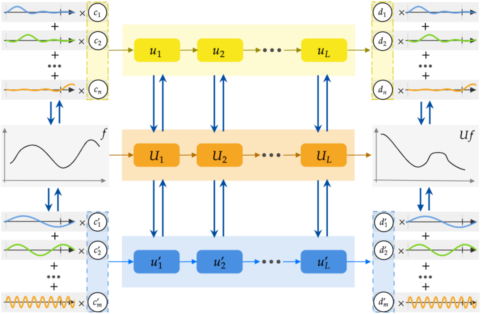
- construct an alias-free discretization u_ℓ of U_ℓ mapping frame coefficients to frame coefficients

$$\begin{array}{ccc} \mathcal{H}_\ell & \xrightarrow{\quad U_\ell \quad} & \mathcal{H}_{\ell+1} \\ \nearrow T_{\Psi_\ell} & & T^*_{\Psi_{\ell+1}} \swarrow \\ \ell^2(I_\ell) & \xrightarrow{\quad u_\ell \quad} & \ell^2(I_{\ell+1}) \end{array}$$

- $\varepsilon(U_\ell, u_\ell, \Psi_\ell, \Psi_{\ell+1}) = 0 \quad \forall \ell = 1, \ldots, L \Longrightarrow \varepsilon(U_\theta, u_\theta) = 0$
- different choices of frames yield different alias-free discretizations

Representation-equivalent Neural Operators





Summary and outlook

- frame theory provides continuous-discrete equivalence and aliasing error for functions
- we generalize this concepts to operators
- define new framework of ReNOs
- Convolutional Neural Operators [Raonić et al., 2023]

Convolutional Neural Operator

- Convolutional Neural Operators [Raonić et al., 2023]
- Layers of a CNO are defined as

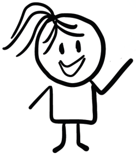
$$U_\ell = \Sigma_\ell \circ \mathcal{K}_\ell, \quad \ell = 1, \dots, L$$

  - $\mathcal{K}_\ell \colon \mathcal{B}_\Omega \to \mathcal{B}_\Omega$ convolution with discrete kernel $\sum_{m,n=-k}^{k} k_{m,n} \delta_{(\frac{m}{2\Omega}, \frac{n}{2\Omega})}$
  - $\Sigma_\ell \colon \mathcal{B}_\Omega \to \mathcal{B}_\Omega$, $\Sigma_\ell = \mathcal{P}_{\mathcal{B}_\Omega} \circ \sigma \circ \mathcal{P}_{\mathcal{B}_{\overline{\Omega}}}$ with $\overline{\Omega} > \Omega$

TU Delft

Thank you for your time!

- F. Bartolucci, E. de Bézenac, B. Raonić, R. Molinaro, S. Mishra, R. Alaifari, **Representation Equivalent Neural Operators: a Framework for Alias-free Operator Learning**, *NeurIPS 2023*
- B. Raonić, R. Molinaro, T. De Ryck, T. Rohner, F. Bartolucci, R. Alaifari, S. Mishra, E. de Bézenac, **Convolutional Neural Operators for robust and accurate learning of PDEs**, *NeurIPS 2023*



Francesca Bartolucci
f.bartolucci@tudelft.nl