

Are Neural Operators Really Neural Operators? Frame Theory Meets Operator Learning

F. Bartolucci and E. de Bézenac and B. Raonic and R. Molinaro and
S. Mishra and R. Alaifari

Research Report No. 2023-21

May 2023

Latest revision: June 2023

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Are Neural Operators Really Neural Operators? Frame Theory Meets Operator Learning

Francesca Bartolucci¹, Emmanuel de Bézenac², Bogdan Raonić^{2,3}, Roberto Molinaro²,
Siddhartha Mishra^{2,3}, and Rima Alaifari^{2,3}

¹Delft University of Technology

²Seminar for Applied Mathematics, ETH Zürich

³ETH AI Center

ABSTRACT. Recently, there has been significant interest in operator learning, i.e. learning mappings between infinite-dimensional function spaces. This has been particularly relevant in the context of learning partial differential equations from data. However, it has been observed that proposed models may not behave as operators when implemented on a computer, questioning the very essence of what *operator learning* should be. We contend that in addition to defining the operator at the continuous level, some form of continuous-discrete equivalence is necessary for an architecture to genuinely learn the underlying operator, rather than just discretizations of it. To this end, we propose to employ *frames*, a concept in applied harmonic analysis and signal processing that gives rise to exact and stable discrete representations of continuous signals. Extending these concepts to operators, we introduce a unifying mathematical framework of *Representation equivalent Neural Operator (ReNO)* to ensure operations at the continuous and discrete level are equivalent. Lack of this equivalence is quantified in terms of aliasing errors. We analyze various existing operator learning architectures to determine whether they fall within this framework, and highlight implications when they fail to do so.

1 Introduction

Operators are mappings between infinite-dimensional function spaces. Prominent examples are *solution operators* for ordinary and partial differential equations (PDEs) [8] which map function space inputs such as initial and boundary data to the function space valued PDE solution. They are also the natural mathematical framework for inverse problems, both in the context of PDEs [10] and in (medical) imaging [2], where the object of interest is the *inverse operator* which maps observables to the underlying material property/image that needs to be inferred or reconstructed.

Given the prohibitive cost of traditional physics based algorithms for approximating operators, particularly those associated with PDEs, increasing attention is being paid in recent years to machine-learning based *operator approximation*. Such techniques for learning operators from

data fall under the genre of *operator learning*. A (by no means comprehensive) list of examples for operator learning architectures include operator networks [5], DeepONets [18], Graph neural operators [16], multipole neural operators [17], PCA-Nets [3], Fourier neural operators (FNO) [15], VIDON [19], spectral neural operators [9], LOCA [11], NOMAD [21] and transformer based operator learning models [4].

However, there is still a lack of clarity on what constitutes *operator learning*? Clearly, an operator learning framework, or a *neural operator* in the nomenclature of [13], should be able to *process functions as inputs and outputs*. On the other hand, functions are continuous objects and in practice, one may not have access to functions either at the input or output level. Instead, one can only access functions and perform computations with them through their *discrete representations* such as point values on a grid, cell averages or in general, coefficients of an underlying basis. Hence, in practice, neural operators have to map discrete inputs into discrete outputs. This leads to a dichotomy, i.e., neural operators need functions as inputs/outputs but only have access to their discrete representations. Consequently, at any finite resolution, possible mismatches between the discretizations and the continuous versions of neural operators can lead to inconsistencies in the underlying function spaces. These inconsistency errors can propagate through the networks and may mar the performance of these algorithms. Moreover, important properties of the underlying operator, such as symmetries and conservation laws, hold at the continuous level. Inconsistent discretizations may not preserve these structural properties of the operator, leading to symmetry breaking etc. with attendant adverse consequences for operator approximation.

Addressing this inconsistency between neural operators and their discretizations is the central point of this article. To this end, we revisit and adapt existing notions of *continuous-discrete equivalence (CDE)* in signal processing, applied harmonic analysis and numerical analysis, by resorting to *frame theory* and the associated questions of sampling and interpolation. Roughly speaking, continuous-discrete equivalence aims at finding mathematical frameworks in which the continuous objects (functions) can be completely (uniquely and stably) recovered from their discrete representations (point evaluations, basis coefficients, etc.) at *any* resolution. Consequently, working with discrete values is tantamount to accessing the underlying continuous object. This equivalence is leveraged to define a *class of neural operators*, for which the discrete input-output representations or their continuous function space realizations are *equivalent*. More concretely, our contributions are

- We provide a novel, very general unifying mathematical formalism to characterize a class of neural operators such that an appropriately defined notion of continuous-discrete equivalence holds for them. These neural operators are termed as *Representation equivalent Neural Operators* or ReNOs. Our definition results in an automatically consistent function space formulation for ReNOs.
- We analyze existing operator learning architectures to find whether they are ReNOs or not.
- Synthetic numerical experiments are presented to illustrate the continuous-discrete equiv-

alence for ReNOs and point out the practical consequences for not respecting it.

2 Continuous-Discrete Equivalences

We start by revisiting concepts on the equivalence of functions and their discrete representations. This is the very essence of *frame theory*, widely used in signal processing and applied harmonic analysis.

Equivalence between functions and their point samples. For simplicity of exposition, we start with univariate functions $f \in L^2(\mathbb{R})$ and the following question: Can a function be uniquely and stably recovered from its values, *sampled* from equispaced grid points $(f(nT))_{n \in \mathbb{Z}}$? The classical *Whittaker-Shannon-Kotel'nikov (WSK) sampling theorem* [22], which lies at the heart of digital-to-analog conversion, answers this question in the affirmative when the underlying function $f \in \mathcal{B}_\Omega$, i.e., it belongs to the *Paley-Wiener space* of *band-limited functions* $\mathcal{B}_\Omega = \{f \in L^2(\mathbb{R}) : \text{supp} \hat{f} \subseteq [-\Omega, \Omega]\}$, for some $\Omega > 0$, \hat{f} the Fourier transform of f and *sampling rate* $1/T \geq 2\Omega$, with 2Ω termed as the *Nyquist rate*. The corresponding reconstruction formula is,

$$f(x) = 2T\Omega \sum_{n \in \mathbb{Z}} f(nT) \text{sinc}(2\Omega(x - nT)), \quad \text{sinc}(x) = \sin(\pi x)/(\pi x). \quad (2.1)$$

We note that the sequence of functions $\{\phi_n(x) = \text{sinc}(2\Omega x - n)\}_{n \in \mathbb{Z}}$ constitutes an *orthonormal basis* for \mathcal{B}_Ω and denote by $\mathcal{P}_{\mathcal{B}_\Omega} : L^2(\mathbb{R}) \rightarrow \mathcal{B}_\Omega$ the *orthogonal projection operator* onto \mathcal{B}_Ω ,

$$\mathcal{P}_{\mathcal{B}_\Omega} f = \sum_{n \in \mathbb{Z}} \langle f, \phi_n \rangle \phi_n = \sum_{n \in \mathbb{Z}} f\left(\frac{n}{2\Omega}\right) \phi_n, \quad (2.2)$$

where the last equality is a consequence of \mathcal{B}_Ω being a reproducing kernel Hilbert space with kernel $\mathcal{K}_\Omega(x, y) = \text{sinc}(2\Omega(x - y))$. Hence, $\mathcal{P}_{\mathcal{B}_\Omega} f$ corresponds to the right-hand side of (2.1) for $T = 1/2\Omega$ and formula (2.1) is exact if and only if $f \in \mathcal{B}_\Omega$, i.e., $f = \mathcal{P}_{\mathcal{B}_\Omega} f \iff f \in \mathcal{B}_\Omega$.

What happens when we sample a function at a sampling rate below the Nyquist rate, i.e. when $1/T < 2\Omega$? Alternatively, what is the effect of approximating a function $f \notin \mathcal{B}_\Omega$ by $\mathcal{P}_{\mathcal{B}_\Omega} f$? Again, sampling theory provides an answer in the form of the *aliasing error*:

Definition 2.1. [1, Section 5] *The aliasing error function $\varepsilon(f)$ and the corresponding aliasing error of $f \in L^2(\mathbb{R})$ for sampling at the rate 2Ω are given by*

$$\varepsilon(f) = f - \mathcal{P}_{\mathcal{B}_\Omega} f, \quad \text{and} \quad \|\varepsilon(f)\|_2 = \|f - \mathcal{P}_{\mathcal{B}_\Omega} f\|_2.$$

If the aliasing error $\varepsilon(f)$ is zero, i.e. if $f \in \mathcal{B}_\Omega$, we say that there is a continuous-discrete equivalence (CDE) between f and its samples $\{f(n/2\Omega)\}_{n \in \mathbb{Z}}$.

Continuous-Discrete Equivalence in Hilbert spaces. Next, we generalize the above described concept of continuous-discrete equivalence to any separable Hilbert space \mathcal{H} , with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. A countable sequence of vectors $\{f_i\}_{i \in I}$ in \mathcal{H} is a *frame* for \mathcal{H} if there exist constants $A, B > 0$ such that for all $f \in \mathcal{H}$

$$A\|f\|^2 \leq \sum_{i \in I} |\langle f, f_i \rangle|^2 \leq B\|f\|^2. \quad (2.3)$$

Clearly, an orthonormal basis for \mathcal{H} is an example of a frame with $A = B = 1$. The bounded operator $T: \ell^2(I) \rightarrow \mathcal{H}$, $T(\{c_i\}_{i \in I}) = \sum_{i \in I} c_i f_i$, is called *synthesis operator* and its adjoint $T^*: \mathcal{H} \rightarrow \ell^2(I)$, $T^*f = \{\langle f, f_i \rangle\}_{i \in I}$, which extracts frame coefficients from an underlying function is called *analysis operator*. By composing T and T^* , we obtain the *frame operator* $S := TT^*$, which is an invertible, self-adjoint and positive operator. Furthermore, the pseudo-inverse of the synthesis operator is given by $T^\dagger: \mathcal{H} \rightarrow \ell^2(I)$, $T^\dagger f = \{\langle f, S^{-1} f_i \rangle\}_{i \in I}$. With these concepts, one can introduce the most prominent result in frame theory [6], the *frame decomposition theorem*, which shows that every element in \mathcal{H} can be *uniquely and stably* reconstructed from its frame coefficients by means of the reconstruction formula

$$f = TT^\dagger f = \sum_{i \in I} \langle f, S^{-1} f_i \rangle f_i = \sum_{i \in I} \langle f, f_i \rangle S^{-1} f_i, \quad (2.4)$$

where the series converge unconditionally. Formula (2.4) is clearly a generalization of reconstruction formula (2.1). However, it is worth pointing out that, in general, the coefficients in (2.4) are not necessarily point samples of the underlying function f , but more general frame coefficients. In general, it may not be possible to access all the frame coefficients to reconstruct a function in a Hilbert space. Instead, just like in the case of reconstructing functions from point samples, one will need to consider approximations to this idealized situation. This is best encapsulated by the notion of a *frame sequence*, i.e., a countable sequence $\{v_i\}_{i \in I} \in \mathcal{H}$ which is a frame for its closed linear span, i.e. for $\overline{\text{span}\{v_i : i \in I\}}$. With this notion, we are in a position to generalize aliasing errors and the CDE, to arbitrary Hilbert spaces. Let \mathcal{H} be a separable Hilbert space and let $\{v_i\}_{i \in I}$ be a frame sequence for \mathcal{H} with $\mathcal{V} = \overline{\text{span}\{v_i : i \in I\}}$ and frame operator $S: \mathcal{V} \rightarrow \mathcal{V}$. Then, the orthogonal projection of \mathcal{H} onto \mathcal{V} is given by $\mathcal{P}_\mathcal{V} f = \sum_{i \in I} \langle f, v_i \rangle S^{-1} v_i$. Thus, formula (2.4) holds and the function f can be uniquely and stably recovered from its frame coefficients if and only if $f \in \mathcal{V}$. If $f \notin \mathcal{V}$, reconstructing f from the corresponding frame coefficients results in an *aliasing error*:

Definition 2.2. The aliasing error function $\varepsilon(f)$ and the resulting aliasing error $\|\varepsilon(f)\|$ of $f \in \mathcal{H}$ for the frame sequence $\{v_i\}_{i \in I} \subseteq \mathcal{V}$ are given by

$$\varepsilon(f) = f - \mathcal{P}_\mathcal{V} f, \quad \|\varepsilon(f)\| = \|f - \mathcal{P}_\mathcal{V} f\|.$$

If the aliasing error $\varepsilon(f)$ is zero, i.e. if $f \in \mathcal{V}$, we say that there is a *continuous-discrete equivalence (CDE)* between f and its frame coefficients $\{\langle f, v_i \rangle\}_{i \in I}$.

3 Representation equivalent Neural Operators (ReNOs).

The concept of neural operators and operator learning [13, 18] can be encapsulated in a pair (U, u) . Here, U is the operator, i.e. the map between infinite-dimensional function spaces, and u is its discretization. As we may have access to different discrete representations associated to the input and output functions in the form of frame coefficients (e.g. point samples evaluated on different grids), u should be able to handle varying inputs and outputs *appropriately*, in a way that is *consistent* with the continuous operator U . Our next aim is to provide a precise mathematical formulation of this issue. To this end, we need to significantly adapt the concepts of CDE and associated aliasing errors, defined for functions in the previous section, to operators.

3.1 Representation equivalent Operators

Let $U: \text{Dom } U \subseteq \mathcal{H} \rightarrow \mathcal{K}$ be an operator between two separable Hilbert spaces, and let $\Psi = \{\psi_i\}_{i \in I}$ and $\Phi = \{\phi_k\}_{k \in K}$ be frame sequences for \mathcal{H} and \mathcal{K} , respectively, with synthesis operators T_Ψ and T_Φ . We denote their closed linear spans by $\mathcal{M}_\Psi := \overline{\text{span}\{\psi_i : i \in I\}}$ and $\mathcal{M}_\Phi := \overline{\text{span}\{\phi_k : k \in K\}}$. We note that by classical frame theory [6], the pseudo-inverses T_Ψ^\dagger and T_Φ^\dagger , initially defined on \mathcal{M}_Ψ and \mathcal{M}_Φ , respectively, can in fact be extended to the entire Hilbert spaces, i.e. $T_\Psi^\dagger: \mathcal{H} \rightarrow \ell^2(I)$ and $T_\Phi^\dagger: \mathcal{K} \rightarrow \ell^2(K)$. Given any mapping $u: \ell^2(I) \rightarrow \ell^2(K)$, we can build the operator $T_\Phi \circ u \circ T_\Psi^\dagger: \mathcal{H} \rightarrow \mathcal{K}$, whose definition clearly depends on the choices of the frame sequences that we make on the continuous level. In other words, any mapping u can be interpreted as a discrete representation of an underlying continuous operator, which in general, may differ from the operator U , that is of interest here. Hence, in analogy to Definitions 2.1 and 2.2, we can define the aliasing error of U relative to the discrete representation u as,

Definition 3.1. *The aliasing error operator $\varepsilon(U, u, \Psi, \Phi)$ of U relative to u and to the choice of frame sequences Ψ and Φ is the operator $\varepsilon(U, u, \Psi, \Phi): \text{Dom } U \subseteq \mathcal{H} \rightarrow \mathcal{K}$ given by*

$$\varepsilon(U, u, \Psi, \Phi) = U - T_\Phi \circ u \circ T_\Psi^\dagger, \quad (3.1)$$

and the corresponding aliasing error is $\|\varepsilon(U, u, \Psi, \Phi)\|$, with $\|\cdot\|$ denoting the operator norm.

By definition, the aliasing error is zero if and only if $U = T_\Phi \circ u \circ T_\Psi^\dagger$, or equivalently if the diagram in Figure 1 commutes, i.e. the black and the blue directed paths in the diagram lead to the same result. If the aliasing error is zero, we say that (U, u, Ψ, Φ) satisfies a *continuous-discrete equivalence (CDE)*, implying that accessing the discrete representation u is exactly the same as accessing the underlying continuous operator U . See **SM D** for an illustrative example of an aliasing error operator.

Equipped with this notion of continuous-discrete equivalence, we can now introduce the concept of *Representation equivalent Operator (ReO)*. To this end, we denote by \mathbf{u} a map defined on pairs (Ψ, Φ) of frame sequences for \mathcal{H} and \mathcal{K} and whose values are mappings from $\ell^2(I)$ into $\ell^2(K)$, i.e. $\mathbf{u}(\Psi, \Phi): \ell^2(I) \rightarrow \ell^2(K)$.

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

Figure 1

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

Figure 2

$$\begin{array}{ccc} \mathcal{H}_{\ell} & \xrightarrow{\mathcal{G}_{\ell}} & \mathcal{H}_{\ell+1} \\ \downarrow T_{\Psi_{\ell}}^{\dagger} & & T_{\Psi_{\ell+1}}^{\dagger} \uparrow \\ \ell^2(I_{\ell}) & \xrightarrow{\mathfrak{u}_{\ell}(\Psi_{\ell}, \Psi_{\ell+1})} & \ell^2(I_{\ell+1}) \end{array}$$

Figure 3

Definition 3.2. We say that (U, \mathfrak{u}) is a Representation equivalent Operator (ReO) if the aliasing error operator $\varepsilon(U, \mathfrak{u}(\Psi, \Phi), \Psi, \Phi)$ is identically zero for every pair (Ψ, Φ) of frame sequences that satisfy

$$\text{Dom } U \subseteq \mathcal{M}_{\Psi} \quad \text{and} \quad \text{Ran } U \subseteq \mathcal{M}_{\Phi}, \quad (3.2)$$

or equivalently, if the diagram in Figure 1 with the choice $u = \mathfrak{u}(\Psi, \Phi)$ commutes.

In other words, we require that the discrete representations $\mathfrak{u}(\Psi, \Phi)$ are all equivalent, meaning that they uniquely determine the same underlying operator U , whenever a continuous-discrete equivalence property holds at the level of the function spaces. If (U, \mathfrak{u}) is a ReO, then any frame sequences $\Psi, \Psi' \subseteq \mathcal{H}, \Phi, \Phi' \subseteq \mathcal{K}$ that satisfy (3.2) result in the following commutative diagram:

$$\begin{array}{ccccc} & & \ell^2(I) & \xrightarrow{\mathfrak{u}(\Psi, \Phi)} & \ell^2(K) \\ & T_{\Psi}^{\dagger} \nearrow & & & \searrow T_{\Phi} \\ \mathcal{H} & \xrightarrow{U} & & & \mathcal{K} \\ & T_{\Psi'}^{\dagger} \searrow & & & \nearrow T_{\Phi'} \\ & & \ell^2(I') & \xrightarrow{\mathfrak{u}(\Psi', \Phi')} & \ell^2(K') \end{array}$$

Thus, there is an equivalence of discrete representations for the same underlying operator U , justifying the name *Representation equivalent Operators*.

Remark 3.3. A possible way to choose the mapping \mathfrak{u} in the above definition is to set,

$$\mathfrak{u}(\Psi, \Phi) = T_{\Phi}^{\dagger} \circ U \circ T_{\Psi}. \quad (3.3)$$

We observe that this definition of $\mathfrak{u}(\Psi, \Phi)$ is such that the diagram in Figure 2 commutes with $u = \mathfrak{u}(\Psi, \Phi)$. Furthermore, the aliasing error operator reads $\varepsilon(U, \mathfrak{u}(\Psi, \Phi), \Psi, \Phi) = U - \mathcal{P}_{\mathcal{M}_{\Phi}} \circ U \circ \mathcal{P}_{\mathcal{M}_{\Psi}}$. This follows by the fact that the compositions $T_{\Psi} \circ T_{\Psi}^{\dagger}$ and $T_{\Phi} \circ T_{\Phi}^{\dagger}$ are the orthogonal projections onto the closed subspaces \mathcal{M}_{Ψ} and \mathcal{M}_{Φ} , respectively. Therefore, the pair (U, \mathfrak{u}) with \mathfrak{u} defined by (3.3) is a ReO. Note that to avoid any aliasing error, the discrete representation of U has to depend on the chosen frame sequences, i.e. inevitably, \mathfrak{u} must depend on Ψ and Φ .

3.2 Representation equivalent Neural Operators

The aim in *operator learning* is to learn non-linear operators $U: \mathcal{H} \rightarrow \mathcal{K}$ between separable Hilbert spaces from input-output function pairs $\{u_i, U(u_i)\}_{i=1}^N$. As with neural networks which

approximate functions mapping finite-dimensional spaces with compositions of *hidden layers*, we will postulate a similar construction for our operator learning architecture. To this end, we will consider the following composition,

$$\mathcal{G} = \mathcal{G}_L \circ \mathcal{G}_{L-1} \circ \dots \circ \mathcal{G}_1 \quad (3.4a)$$

of linear and non-linear operators

$$\mathcal{G}_\ell: \mathcal{H}_\ell \rightarrow \mathcal{H}_{\ell+1}, \quad \ell = 1, \dots, L, \quad (3.4b)$$

between separable Hilbert spaces. For each $1 \leq \ell \leq L$, \mathcal{G}_ℓ is termed a *hidden layer*.

Note that the form (3.4) of the operator \mathcal{G} is very general. A particular example is provided by the neural operators (Eqn. 5) of [13], where each hidden layer is of the form $\mathcal{G}_\ell = \sigma(W_\ell + K_\ell + b_\ell)$, with W_ℓ a local (pointwise) linear map (matrix), K_ℓ a non-local linear operator, for instance a kernel-integral operator and b_ℓ a bias function. The non-linearity is introduced through the activation function σ which acts pointwise. However, we prefer to keep the form (3.4) as it can represent much more general operators, for instance when the non-linearity acts in a non-local manner.

In practice, it is not possible to directly work with the continuous form (3.4), whether at the level of accessing inputs/outputs or performing computations on a digital computer. Rather, we have to rely on a discretization of (3.4). Given the notion of *Representation equivalent Operators* above, wherein the continuous form operators can be uniquely and stably recovered from its discretizations, we particularize this concept to neural operators. To this end, we consider a finite sequence of separable Hilbert spaces $\{\mathcal{H}_\ell\}_{\ell=1}^{L+1}$ in (3.4) and for every $\ell = 1, \dots, L+1$, we denote by $\Psi_\ell = \{\psi_{\ell,i}\}_{i \in I_\ell}$ a *frame sequence* for \mathcal{H}_ℓ . The choice of frame sequences for each \mathcal{H}_ℓ corresponds to the choice of accessible discrete representations of functions in the underlying function space \mathcal{H}_ℓ . For instance, if $\mathcal{H}_\ell = \mathcal{B}_\Omega$, the space of bandlimited functions with bandwidth $\Omega > 0$, and we can access the pointwise evaluations of any function $f \in \mathcal{H}_\ell$ on the uniform grid $\{\frac{n}{2\Omega}\}_{n \in \mathbb{Z}}$, then Ψ_ℓ may be chosen as the orthonormal basis $\{\text{sinc}(2\Omega x - n)\}_{n \in \mathbb{Z}}$.

Now let $(\mathcal{G}, \mathfrak{g})$ be a pair of the form

$$\mathcal{G} = \mathcal{G}_L \circ \mathcal{G}_{L-1} \circ \dots \circ \mathcal{G}_1, \quad \mathfrak{g} = \mathfrak{g}_L \circ \mathfrak{g}_{L-1} \circ \dots \circ \mathfrak{g}_1, \quad (3.5)$$

where, for every $\ell = 1, \dots, L$, $\mathcal{G}_\ell: \mathcal{H}_\ell \rightarrow \mathcal{H}_{\ell+1}$ is a (possibly) non-linear operator and $\mathfrak{g}_\ell(\Psi_\ell, \Psi_{\ell+1}): \text{Ran } T_{\Psi_\ell}^\dagger \subseteq \ell^2(I_\ell) \rightarrow \text{Ran } T_{\Psi_{\ell+1}}^\dagger \subseteq \ell^2(I_{\ell+1})$ for each choice of frame sequences $\Psi_\ell \subseteq \mathcal{H}_\ell, \Psi_{\ell+1} \subseteq \mathcal{H}_{\ell+1}$.

Definition 3.4. *The pair $(\mathcal{G}, \mathfrak{g})$ is called a Representation equivalent Neural Operator (ReNO) if it is a Representation equivalent Operator at each hidden layer, meaning that for every $\ell = 1, \dots, L$ the aliasing error operator $\varepsilon(\mathcal{G}_\ell, \mathfrak{g}_\ell(\Psi_\ell, \Psi_{\ell+1}), \Psi_\ell, \Psi_{\ell+1})$ is identically zero for every pair $(\Psi_\ell, \Psi_{\ell+1})$ of frame sequences that satisfy $\text{Dom } \mathcal{G}_\ell \subseteq \mathcal{M}_{\Psi_\ell}$ and $\text{Ran } \mathcal{G}_\ell \subseteq \mathcal{M}_{\Psi_{\ell+1}}$. Or equivalently, the diagram in Figure 3 is commutative for every ℓ .*

Remark 3.5. *If the aliasing error operator $\varepsilon(\mathcal{G}_\ell, \mathbf{g}_\ell(\Psi_\ell, \Psi_{\ell+1}), \Psi_\ell, \Psi_{\ell+1})$ is identically zero (as required in Definition 3.4), then the assumption that $\mathbf{g}_\ell(\Psi_\ell, \Psi_{\ell+1})$ maps $\text{Ran } T_{\Psi_\ell^\dagger}$ into $\text{Ran } T_{\Psi_{\ell+1}^\dagger}$ implies that $\mathbf{g}_\ell(\Psi_\ell, \Psi_{\ell+1})$ is of the form (3.3), i.e. $\mathbf{g}_\ell(\Psi_\ell, \Psi_{\ell+1}) = T_{\Psi_{\ell+1}^\dagger}^\dagger \circ \mathcal{G}_\ell \circ T_{\Psi_\ell}$. This readily shows that there is only one recipe for constructing ReNOs (see SM B for the proof). In particular, every map \mathbf{g}_ℓ satisfies*

$$\mathbf{g}_\ell(\Psi'_\ell, \Psi'_{\ell+1}) = T_{\Psi'_{\ell+1}^\dagger}^\dagger \circ T_{\Psi_{\ell+1}} \circ \mathbf{g}_\ell(\Psi_\ell, \Psi_{\ell+1}) \circ T_{\Psi_\ell}^\dagger \circ T_{\Psi'_\ell}, \quad (3.6)$$

whenever the frame sequences satisfy (3.2). Thus, Eqn. (3.6) characterizes the discretized versions of ReNO under changes of frame sequences.

Definition 3.4 requires that at each hidden layer the underlying operator satisfies a continuous-discrete equivalence property and accessing a discretized version is exactly the same as accessing the underlying continuous operator whenever the frame sequences fulfill (3.2) at each layer. Requiring this CDE at each hidden layer suffices to enforce it at the level of the operator (3.4) itself:

Proposition 3.6. *If the pair $(\mathcal{G}, \mathbf{g})$ is a ReNO, then the operator \mathcal{G} in (3.4) is uniquely determined by its discretized version \mathbf{g} in (3.5) since the aliasing error (3.1) is identically zero.*

As the proof of Proposition 3.6, presented in SM A, also shows, if each hidden layer in the operator (3.4) has an aliasing error (3.1), then these errors may propagate through the network and increase with increasing number of layers.

Remark 3.7. *The goal in operator learning is to learn an operator $U: \mathcal{H} \rightarrow \mathcal{K}$ between separable Hilbert spaces \mathcal{H} and \mathcal{K} from a discrete representation of its sampled values $\{u_i, U(u_i)\}_{i=1}^N \in \mathcal{H} \times \mathcal{K}$. The above discussion indicates it is crucial to maintain an exact correspondence between the continuous and the discrete representations in the training set $\{u_i, U(u_i)\}_{i=1}^N$ as well. In other words, no aliasing should be present in order to learn the target operator correctly. Mathematically, this can be formulated in terms of accessing the discrete representation $\{T_\Psi^\dagger(u_i), T_\Phi^\dagger(U(u_i))\}_{i=1}^N$ of $\{u_i, U(u_i)\}_{i=1}^N$ for underlying frame sequences $\Psi \subseteq \mathcal{H}$ and $\Phi \subseteq \mathcal{K}$ that satisfy (3.2).*

4 Examples

Equipped with our definition of Representation equivalent Neural Operators (ReNOs), we analyze some existing operator learning architectures to ascertain whether they are neural operators or not.

Convolutional Neural Networks. Classical convolutional neural networks (CNN) are based on the convolutional layer K , involving a discrete kernel $k \in \mathbb{R}^{2s+1}$ and a discrete input c :

$$K(c)[m] = (k * c)[m] = \sum_{i=-s}^s c[m-i]k[i].$$

We can then analyze this layer using our framework to ask whether this operation can be associated to some underlying continuous operator. Intuitively, if this is the case, the computations conducted on different discretizations effectively representing the input should be consistent; in the contrary case, no associated continuous operator exists and the convolutional operation is not a ReO.

Consider the case where the discrete input c corresponds to pointwise evaluation on a grid of some underlying bandlimited function $f \in \mathcal{B}_\Omega$, for example $c[n] = f\left(\frac{n}{2\Omega}\right)$, $n \in \mathbb{Z}$ with associated orthonormal basis $\Psi = \{\text{sinc}(2\Omega x - n)\}_{n \in \mathbb{Z}}$. Consider now an alternate representation of f , point samples of a grid twice as fine: $d[n] = f\left(\frac{n}{4\Omega}\right)$, with $\Psi' = \{\text{sinc}(4\Omega x - n)\}_{n \in \mathbb{Z}}$ as basis. Clearly, even though discrete inputs agree, i.e. $c[n] = d[2n]$, this is no longer true for the outputs, $(k * c)[n] \neq (k * d)[2n]$. This in turn implies that there exist frame sequences Ψ, Ψ' such that:

$$T_\Psi \circ K_\Phi(c) \circ T_\Psi^\dagger \neq T_{\Psi'} \circ K_{\Phi'}(d) \circ T_{\Psi'}^\dagger, \quad (4.1)$$

thereby defying the representation equivalence property of ReOs. As a consequence, CNNs are not ReNOs, in the sense of Definition 3.4. This fact is also corroborated with the experimental analysis, Section 5.

Fourier Neural Operators (FNO). FNOs are defined in [15] in terms of layers, which are either lifting or projection layers or *Fourier layers*. As lifting and projection layers do not change the underlying spatial structure of the input function, but only act on the channel width, these linear layers will satisfy Definition 3.4 of ReNO here. Hence, we focus on the Fourier layer of the form,

$$v_{\ell+1}(x) = \sigma(A_\ell v_\ell(x) + B_\ell(x) + \mathcal{K}v_\ell(x)), \quad (4.2)$$

with the Fourier operator given by

$$\mathcal{K}v = \mathcal{F}^{-1}(R \odot \mathcal{F})(v).$$

Here, $\mathcal{F}, \mathcal{F}^{-1}$ are the Fourier and Inverse Fourier transforms.

For simplicity of exposition, we set $A_\ell = B_\ell \equiv 0$ and focus on investigating whether the Fourier layer (4.2) satisfies the requirements of a ReNO. Following [15], the discrete form of the Fourier layer is given by $\sigma(kv)$, with $kv = F^{-1}(R \odot F(v))$, where F, F^{-1} denote the discrete Fourier transform (DFT) and its inverse.

In **SM C**, we show that the convolution in Fourier space operation for the FNO layer (4.2) satisfies the requirements of a ReNO. However the pointwise activation function $\sigma(f)$, applied to a bandlimited input $f \in \mathcal{B}_\Omega$ will not necessarily respect the bandlimits, i.e., $\sigma(f) \notin \mathcal{B}_\Omega$. In fact, with popular choices of activation functions such as ReLU, $\sigma(f) \notin \mathcal{B}_\omega$, for any $\omega > 0$ (see **SM F** for numerical illustrations). Thus, the Fourier layer operator (4.2) may not respect the continuous-discrete equivalence and can lead to aliasing errors, a fact already identified in [9]. Hence, FNO *may not be a ReNO in the sense of Definition 3.4*.

Spectral Neural Operators (SNO). Introduced in [9], this architecture is defined as follows. Let $K > 0$ and denote by

$$\mathcal{P}_K = \left\{ g(x) = \sum_{k=-K}^K c_k e^{i\pi k x} : (c_k)_{k=-K}^K \in \mathbb{C}^{2K+1} \right\},$$

the space of 2-periodic signals bandlimited to πK . Clearly, $\Psi_K = \{e^{i\pi k \cdot}\}_{k=-K}^K$ constitutes an orthonormal basis for \mathcal{P}_K , and the corresponding synthesis operator $T_{\Psi_K} : \mathbb{C}^{2K+1} \rightarrow \mathcal{P}_K$ and analysis operator $T_{\Psi_K}^* : \mathcal{P}_K \rightarrow \mathbb{C}^{2K+1}$ are given by

$$T_{\Psi_K}((c_k)_{k=-K}^K) = \sum_{k=-K}^K c_k e^{i\pi k \cdot}, \quad T_{\Psi_K}^* f = \{\langle f, e^{i\pi k \cdot} \rangle\}_{k=-K}^K.$$

A spectral neural operator is defined as the compositional mapping $T_{\Psi_{K'}} \circ \mathcal{N} \circ T_{\Psi_K}^*$, where $\mathcal{N} : \mathbb{C}^{2K+1} \rightarrow \mathbb{C}^{2K'+1}$ is an ordinary feedforward neural network with activation function σ ,

$$\mathcal{N}(x) = W^{(L+1)} \sigma(W^{(L)} \dots \sigma(W^{(2)} \sigma(W^{(1)} x - b^{(1)}) - b^{(2)}) \dots - b^{(L)}) - b^{(L+1)} \quad (4.3)$$

for some weights $W^{(\ell)}$ and biases $b^{(\ell)}$, $\ell = 1, \dots, L+1$. It is straightforward to see that this architecture corresponds to the commutative diagram,

$$\begin{array}{ccccccc} \mathcal{P}_K & \xrightarrow{T_{\Psi_K}^*} & \mathbb{C}^{2K+1} & \xrightarrow{\mathcal{N}} & \mathbb{C}^{2K'+1} & \xrightarrow{T_{\Psi_{K'}}} & \mathcal{P}_{K'} \\ T_{\Psi_K} \uparrow & & \uparrow \text{Id} & & \uparrow \text{Id} & & T_{\Psi_{K'}} \uparrow \\ \mathbb{C}^{2K+1} & \xrightarrow{\text{Id}} & \mathbb{C}^{2K+1} & \xrightarrow{\mathcal{N}} & \mathbb{C}^{2K'+1} & \xrightarrow{\text{Id}} & \mathbb{C}^{2K'+1} \end{array}$$

A discretized version of spectral neural operators simply corresponds to an ordinary feedforward neural network mapping Fourier coefficients to Fourier coefficients. We conclude that for SNOs to be ReNOs with respect to the function spaces $\mathcal{P}_K, \mathcal{P}'_K$, we have to enforce that for more general choices of frame sequences, Remark 3.5 is satisfied. Moreover, the architecture of SNOs can be generalized with respect to any frames in finite-dimensional inner-product spaces. Under some appropriate condition on the output spanned by DeepONets [18], they also fall in this class.

5 Experimental Analysis

An empirical analysis was conducted to demonstrate the pertinence of the framework presented in Section 3, and to support the theoretical claims made in Section 4. This experiment consists in training each of the models presented of Section 4 on a fixed discretization, and evaluating how they behave when alternate discretizations are considered.

Regression to some given target operator $Q : \mathcal{H} \rightarrow \mathcal{H}$ is considered, where \mathcal{H} is the space of bandlimited and periodic functions. The precise construction of the target operator is deferred to SM F, as well as other experimental details. Frames (Ψ, Φ) for \mathcal{H} are considered, discrete

representations of the training data are derived from these, which are used to train a discrete operator $u : \ell^2(I) \rightarrow \ell^2(K)$.

After training, we consider alternate discretizations derived from frame sequences (Ψ', Φ') , with their associated mapping $u' : \ell^2(I') \rightarrow \ell^2(K')$. Discrepancy between the two is then calculated by computing each mapping on their given discretization, and comparing them by expressing u' in the basis associated to u . This idea is made formal introducing the discrete aliasing error map $\varepsilon(u, u') : \ell^2(I) \rightarrow \ell^2(K)$, defined as:

$$\varepsilon(u, u') = u - T_{\Phi}^{\dagger} \circ T_{\Phi'} \circ u' \circ T_{\Psi'}^{\dagger} \circ T_{\Psi} \tag{5.1}$$

Remark 5.1. *In the case of an orthogonal basis, the discrete aliasing error $\varepsilon(u, u')$ precisely corresponds to the discretization of the aliasing error operator, as $\varepsilon(u, u') = T_{\Phi}^{\dagger} \circ \varepsilon(U, u', \Psi', \Phi') \circ T_{\Psi}$ with $U = T_{\Phi} \circ u \circ T_{\Psi}^{\dagger}$. This implies that if we consider ReNOs, $\varepsilon(u, u') = 0$ as long as (Ψ', Φ') correspond to frame sequences satisfying the spanning condition in (3.2).*

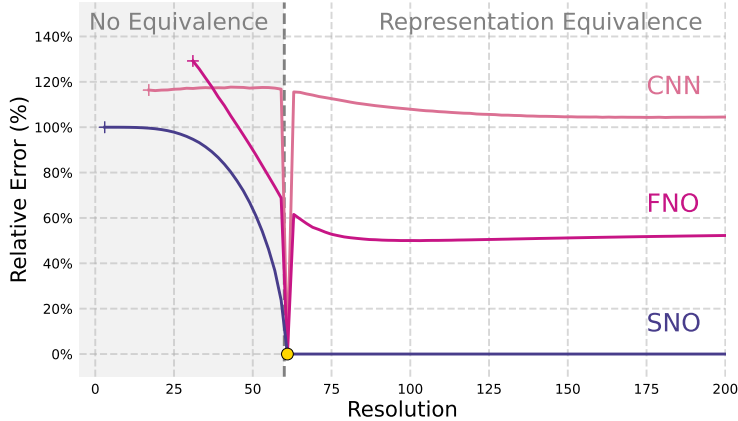


Figure 4: Representation equivalence analysis is conducted by training three classical architectures on a fixed resolution and examining their performance when input resolution is varied. The resolution equivalence zone, located on the right-hand side, denotes the region where discrete representations have an associated frame, while the left-hand side represents the area where this is no longer the case. As predicted by the theory, CNN and FNO are not representation equivalent, while SNO is a ReNO in the representation equivalence region, but failing to do so outside of it.

The results of this experiment are presented in Figure 4. We test 3 models in this experiment, namely CNN, FNO and SNO. All the models are trained on a given resolution (61^2) and evaluated on varying resolutions. The results in Figure 4 clearly show that as predicted by our theory, neither CNN nor FNO are representation equivalent, in the sense of Definition 3.4 and changing the resolution, which amounts to a change of frame, does not keep the operator invariant and causes aliasing errors. On the other hand, as predicted by the theory, SNO is a

ReNO as long as the conditions of Definition 3.4 and Remark 3.5 hold. When these conditions no longer hold, SNO also generates aliasing errors. Thus, this experiment clearly demonstrates the practical implications of not respecting representation or continuous-discrete equivalence.

6 Discussion

Summary. Although a variety of architectures for operator learning are available since the last couple of years, there is still a lack of clarity about what exactly constitutes *operator learning*. Everyone would agree that an operator learning architecture should have functions as inputs and outputs. However, in practice, one does not necessarily have access to functions, either at the input or output level. Rather the access is limited to some form of *discrete representations* of the underlying functions, for instance, point values, cell averages, coefficients with respect to a basis etc. Moreover, one can only perform computations with discrete objects on digital computers. Hence, given the necessity of having to work with discrete representations of functions, it is essential to enforce some relationship between the continuous and discrete representations of the underlying functions or in other words, demand consistency in function space.

Unlike in [13], where the authors advanced a form of *asymptotic consistency* in terms of discretization parameters, we go further to require structure-preserving *continuous-discrete equivalence (CDE)*. To this end, we leverage the well-established concept of CDE for functions in Hilbert spaces in terms of frame theory. The main point about this CDE is the fact that functions, belonging to suitable function spaces, can be uniquely and stably reconstructed, from their frame coefficients. A failure to enforce this CDE results in the so-called aliasing errors that quantitatively measure function space inconsistencies.

We extend these notions of CDE and aliasing errors to operators here. These notions of CDE for operators are then used to define *Representation equivalent Neural Operators (ReNOs)*, see Definition 3.4. Our definition amounts to requiring the CDE for each layer of the approximating operator. This will ensure that there is no aliasing error when the composite operator is considered. Moreover, our framework automatically implies consistency in function spaces and provides a recipe for deriving ReNOs in terms of changing the underlying frames. We also employ our framework to analyze whether or not existing operator learning architectures are ReNOs and also corroborate our results through experiments.

Related Work. Our current paper relies heavily on structure preservation from classical numerical analysis [20] and references therein and CDE concepts from signal processing and applied harmonic analysis [22] and references therein. Among the emerging literature on operator learning, [13] was among the first papers to attempt a unifying framework that encompasses many operator learning architectures and codify them through a particular definition of neural operators. We take an analogous route here but with our main point of departure from [13] being that unlike their notion of asymptotic consistency, we require *systematic consistency in function space* by enforcing the representation equivalence for the underlying operator at each

layer. Another relevant work is [9] where the authors flag the issue of possible aliasing errors with specific operator learning architectures. We significantly expand on the approach of [9] by providing a rigorous and very general definition for aliasing errors. Finally, our definition of Representation equivalent Neural Operators, leveraging CDEs for operators, is analogous to a similar approach for *computing with functions*, rather than discrete representations, which is the cornerstone of the *Chebfun* project [7] and references therein.

Limitations and Extensions. Our aim in this paper was to tackle a fundamental question of what defines a *neural operator*? We have addressed this question here and shown that enforcing some form of CDE is needed for the architecture to genuinely learn the underlying operator rather than just a discrete representation of it. What we have not addressed here are quantitative measures of the error associated with a *Representation equivalent Neural Operator*, introduced in Definition 3.4, in approximating the underlying operator. This is a much broader question than what is addressed here, since sources of error, other than aliasing errors, such as *approximation, training and generalization* errors also contribute to the total error (see [14, 12, 13]).

One interesting direction for further analysis would involve exploring operators and their discretized counterparts that adhere to a relaxed form of CDE. In our formulation, this relaxation could be achieved by allowing the aliasing error to be bounded by a specified value instead of mandating it to be exactly zero.

Supplementary Material for:
 Are Neural Operators Really Neural Operators?
 Frame Theory Meets Operators Learning

A Proof of Proposition 3.6

Proof. We show the proof of Proposition 3.6 for a two layer Representation equivalent Neural Operator (3.5). The proof generalizes to the case of $\ell > 2$ layers. We want to show that the aliasing error $\varepsilon(\mathcal{G}_2 \circ \mathcal{G}_1, \mathfrak{g}(\Psi_2, \Psi_3) \circ \mathfrak{g}(\Psi_1, \Psi_2), \Psi_1, \Psi_3)$ is identically zero whenever the frame sequences satisfy (3.2), or equivalently whenever the diagram

$$\begin{array}{ccccc} \mathcal{H}_1 & \xrightarrow{\mathcal{G}_1} & \mathcal{H}_2 & \xrightarrow{\mathcal{G}_2} & \mathcal{H}_3 \\ \downarrow T_{\Psi_1}^\dagger & & T_{\Psi_2} \uparrow T_{\Psi_2}^\dagger & & T_{\Psi_3} \uparrow \\ \ell^2(I_1) & \xrightarrow{\mathfrak{g}(\Psi_1, \Psi_2)} & \ell^2(I_2) & \xrightarrow{\mathfrak{g}(\Psi_2, \Psi_3)} & \ell^2(I_3) \end{array}$$

is commutative. We directly compute

$$\begin{aligned} \mathcal{G}_2 \circ \mathcal{G}_1 &= (T_{\Psi_3} \circ \mathfrak{g}(\Psi_2, \Psi_3) \circ T_{\Psi_2}^\dagger) \circ (T_{\Psi_2} \circ \mathfrak{g}(\Psi_1, \Psi_2) \circ T_{\Psi_1}^\dagger) \\ &= T_{\Psi_3} \circ \mathfrak{g}(\Psi_2, \Psi_3) \circ \mathfrak{g}(\Psi_1, \Psi_2) \circ T_{\Psi_1}^\dagger. \end{aligned}$$

The first equality simply follows by the definition of ReNO. The last equality can be seen as follows: first, $\mathcal{M}_{\Psi_2} = \text{Ran}(T_{\Psi_2} T_{\Psi_2}^*) \subseteq \text{Ran}(T_{\Psi_2}) \subseteq \mathcal{M}_{\Psi_2}$, so that $\text{Ran}(T_{\Psi_2})$ is closed. This, in combination with Lemma 2.5.2 in [6], implies that $\text{Ran}(T_{\Psi_2}^\dagger)$ is also closed. Hence, $T_{\Psi_2}^\dagger \circ T_{\Psi_2}$ is the orthogonal projection onto $(\text{Ker}(T_{\Psi_2}))^\perp = \text{Ran}(T_{\Psi_2}^\dagger)$ and, by assumption, $\text{Ran}(\mathfrak{g}(\Psi_1, \Psi_2) \circ T_{\Psi_1}^\dagger) \subseteq \text{Ran}(T_{\Psi_2}^\dagger)$. As a consequence, the aliasing error operator (3.1) of $\mathcal{G}_2 \circ \mathcal{G}_1$ for the discretized version $\mathfrak{g}(\Psi_2, \Psi_3) \circ \mathfrak{g}(\Psi_1, \Psi_2)$ is identically zero, which proves the hypothesis. \square

B Proof of Remark 3.5

We keep the notation as in Section 3.2. Let $(\mathcal{G}, \mathfrak{g})$ be a pair of the form

$$\mathcal{G} = \mathcal{G}_L \circ \mathcal{G}_{L-1} \circ \dots \circ \mathcal{G}_1, \quad \mathfrak{g} = \mathfrak{g}_L \circ \mathfrak{g}_{L-1} \circ \dots \circ \mathfrak{g}_1,$$

where, for every $\ell = 1, \dots, L$, $\mathcal{G}_\ell: \mathcal{H}_\ell \rightarrow \mathcal{H}_{\ell+1}$ is a (possibly) non-linear operator and $\mathfrak{g}_\ell(\Psi_\ell, \Psi_{\ell+1}): \text{Ran } T_{\Psi_\ell}^\dagger \subseteq \ell^2(I_\ell) \rightarrow \text{Ran } T_{\Psi_{\ell+1}}^\dagger \subseteq \ell^2(I_{\ell+1})$ for each choice of frame sequences $\Psi_\ell \subseteq \mathcal{H}_\ell, \Psi_{\ell+1} \subseteq \mathcal{H}_{\ell+1}$. By definition, if $(\mathcal{G}, \mathfrak{g})$ is a ReNO, then

$$\mathcal{G}_\ell = T_{\Psi_{\ell+1}} \circ \mathfrak{g}_\ell(\Psi_\ell, \Psi_{\ell+1}) \circ T_{\Psi_\ell}^\dagger \tag{B.1}$$

for every pair $(\Psi_\ell, \Psi_{\ell+1})$ of frame sequences that satisfy $\text{Dom } \mathcal{G}_\ell \subseteq \mathcal{M}_{\Psi_\ell}$ and $\text{Ran } \mathcal{G}_\ell \subseteq \mathcal{M}_{\Psi_{\ell+1}}$. By equation (B.1) we readily obtain

$$T_{\Psi_{\ell+1}}^\dagger \circ \mathcal{G}_\ell \circ T_{\Psi_\ell} = T_{\Psi_{\ell+1}}^\dagger \circ T_{\Psi_{\ell+1}} \circ \mathfrak{g}_\ell(\Psi_\ell, \Psi_{\ell+1}) \circ T_{\Psi_\ell}^\dagger \circ T_{\Psi_\ell} = \mathfrak{g}_\ell(\Psi_\ell, \Psi_{\ell+1}),$$

where the last equality follows by the fact that $T_{\Psi_\ell}^\dagger \circ T_{\Psi_\ell}$ is the orthogonal projection onto $(\text{Ker}(T_{\Psi_\ell}))^\perp = \text{Ran}(T_{\Psi_\ell}^\dagger)$ and, by assumption, $\mathfrak{g}(\Psi_\ell, \Psi_{\ell+1})$ maps $\text{Ran } T_{\Psi_\ell}^\dagger$ into $\text{Ran } T_{\Psi_{\ell+1}}^\dagger$. This concludes the proof of Remark 3.5.

C Fourier layer in FNOs

We focus here on the Fourier layer of FNOs, i.e.

$$\mathcal{K}v = \mathcal{F}^{-1}(R \odot \mathcal{F})(v), \quad (\text{C.1})$$

where $\mathcal{F}, \mathcal{F}^{-1}$ denote the Fourier transform and its inverse, and where R denotes a low-pass filter. In [15], the authors define the Fourier layer on the space $L^2(\mathbb{T})$ of 2-periodic functions and, with slight abuse of notation, they refer to the mappings $\mathcal{F}: L^2(\mathbb{T}) \rightarrow \ell^2(\mathbb{Z})$ and $\mathcal{F}^{-1}: \ell^2(\mathbb{Z}) \rightarrow L^2(\mathbb{T})$,

$$\mathcal{F}w(k) = \langle w, e^{i\pi kx} \rangle, \quad \mathcal{F}^{-1}(\{W_k\}_{k \in \mathbb{Z}}) = \sum_{k \in \mathbb{Z}} W_k e^{i\pi kx},$$

as the Fourier transform and the inverse Fourier transform. Furthermore, the authors define the discrete version of (C.1) as $F^{-1}(R \odot F)$, where F, F^{-1} denote the discrete Fourier transform (DFT) and its inverse, and where they assume to have access only to point-wise evaluations of the input and output functions. However, *the space of 2-periodic functions is too large to allow for any form of continuous-discrete equivalence (CDE)* when the input v and the output $\mathcal{K}v$ are represented by their point samples. Consequently, here we consider smaller subspaces of $L^2(\mathbb{T})$ which allow for CDEs. More precisely, we are able to show that FNO Fourier layers can be realized as Representation equivalent Operators (crf. Definition 3.2) between bandlimited and periodic functions. Let $K > 0$ and let \mathcal{P}_K be the space of bandlimited 2-periodic functions

$$\mathcal{P}_K = \left\{ w(x) = \sum_{k=-K}^K W_k e^{i\pi kx} : \{W_k\}_{k=-K}^K \in \mathbb{C}^{2K+1} \right\}.$$

Every function $w \in \mathcal{P}_K$ can be uniquely represented by its Fourier coefficients $\{W_k\}_{k=-K}^K$ as well as by its samples $\{w(\frac{k}{2K+1})\}_{k=-K}^K$, see [23, Section 5.5.2]. Indeed, the latter ones are the coefficients of w with respect to the orthonormal basis

$$\Psi_K = \left\{ \frac{1}{\sqrt{2(2K+1)}} d\left(\cdot - \frac{2k}{2K+1}\right) \right\}_{k=0}^{2K}, \quad (\text{C.2})$$

where d denotes the Dirichlet kernel of order K and period 2, defined as

$$d(t) = \sum_{k=-K}^K e^{i\pi kt}.$$

Furthermore, the DFT $\{\widehat{W}_k\}_{k=-K}^K$ of the sample sequence $\{w(\frac{k}{2K+1})\}_{k=-K}^K$ is related to the Fourier coefficients of w via the equation

$$\widehat{W}_k = (2K+1)W_k, \quad k = -K, \dots, K,$$

and we refer to [23, Section 5.5.2] for its proof. Thus, this yields the commutative diagram

$$\begin{array}{ccc} \mathcal{P}_K & \xrightarrow{\mathcal{F}} & \mathbb{C}^{2K+1}, \\ \downarrow T_{\Psi_K}^\dagger & & \uparrow \text{Id} \\ \mathbb{C}^{2K+1} & \xrightarrow{(2K+1)\cdot\mathcal{F}} & \mathbb{C}^{2K+1} \end{array}$$

where $T_{\Psi_K}^\dagger : \mathcal{P}_K \rightarrow \mathbb{C}^{2K+1}$ denotes the analysis operator associated to the basis (C.2). Analogously, we can build the commutative diagram

$$\begin{array}{ccc} \mathbb{C}^{2K'+1} & \xrightarrow{\mathcal{F}^{-1}} & \mathcal{P}_{K'} \\ \downarrow \text{Id} & & \uparrow T_{\Psi_{K'}} \\ \mathbb{C}^{2K'+1} & \xrightarrow{\frac{1}{(2K'+1)}\cdot\mathcal{F}^{-1}} & \mathbb{C}^{2K'+1} \end{array}$$

where $T_{\Psi_{K'}} : \mathbb{C}^{2K'+1} \rightarrow \mathcal{P}_{K'}$ denotes the synthesis operator associated to the basis (C.2) with $K = K'$. Then, $R = \{R_k\}_{k=-K'}^{K'}$, with $K' \leq K$, denotes the Fourier coefficients of a 2-periodic function, and the mapping $R \odot \mathcal{F} : \mathcal{P}_K \rightarrow \mathbb{C}^{2K'+1}$ is defined as

$$(R \odot \mathcal{F}w)(k) = R_k W_k, \quad k = -K', \dots, K'.$$

Therefore, by definition, $R \odot \mathcal{F}$ yields a continuous-discrete equivalence operation. Overall, we get the commutative diagram

$$\begin{array}{ccccccc} \mathcal{P}_K & \xrightarrow{\mathcal{F}} & \mathbb{C}^{2K+1} & \xrightarrow{R \odot} & \mathbb{C}^{2K'+1} & \xrightarrow{\mathcal{F}^{-1}} & \mathcal{P}_{K'} \\ \downarrow T_{\Psi_K}^\dagger & & \uparrow \text{Id} & & \uparrow \text{Id} & & \uparrow T_{\Psi_{K'}} \\ \mathbb{C}^{2K+1} & \xrightarrow{(2K+1)\cdot\mathcal{F}} & \mathbb{C}^{2K+1} & \xrightarrow{R \odot} & \mathbb{C}^{2K'+1} & \xrightarrow{\frac{1}{(2K'+1)}\cdot\mathcal{F}^{-1}} & \mathbb{C}^{2K'+1} \end{array}$$

which shows that the discretization of the Fourier layer C.1, the blue path in the above commutative diagram, is defined via Equation (3.3). As a consequence, the Fourier layer C.1, regarded as an operator from \mathcal{P}_K into $\mathcal{P}_{K'}$ satisfies the requirements of a Representation equivalent Operator (crf. Definition 3.2). However, as pointed out in 4, the pointwise activation function applied to a bandlimited input will not necessarily respect the bandwidth. In fact, with popular choices of activation functions such as ReLU, $\sigma(f) \notin \mathcal{P}_K$, for any $K \in \mathbb{N}$ (see also **SM F** for numerical illustrations). Thus, the FNO layer

$$\sigma(\mathcal{K}v) = \sigma(\mathcal{F}^{-1}(R \odot \mathcal{F})(v))$$

may not respect the continuous-discrete equivalence and can lead to aliasing errors, a fact already identified in [9]. Hence, FNOs *may not be ReNOs in the sense of Definition 3.4*.

D An Illustrative Example of Aliasing in the context of Operator Learning

Let $\mathcal{B}_\Omega \subset L^2(\mathbb{R})$ denote the space of bandlimited functions with bandwidth $\Omega > 0$. We consider the non-linear operator $U : \mathcal{B}_\Omega \rightarrow \mathcal{B}_{2\Omega}$ defined by $U(f) = |f|^2$. Suppose we take

$\Psi = \{\text{sinc}(2\Omega x - n)\}_{n \in \mathbb{Z}}$ as an orthonormal basis for \mathcal{B}_Ω and the same sequence $\Phi = \Psi$ as a frame sequence for $\mathcal{B}_{2\Omega}$, and consider the associated discretized version $\mathbf{u}(\Psi, \Phi)$ of U according to (3.3). This is equivalent to assuming that we only have access to pointwise evaluations of input and output functions on the grid $\{\frac{n}{2\Omega}\}_{n \in \mathbb{Z}}$. Such a choice for the discrete representation of the functions causes aliasing in the sense of Definition 3.1, since for every $f \in \mathcal{B}_\Omega$ such that $U(f) \in \mathcal{B}_{2\Omega} \setminus \mathcal{B}_\Omega$, we have

$$\varepsilon(U, \mathbf{u}(\Psi, \Phi), \Psi, \Phi)(f) = U(f) - \mathcal{P}_{\mathcal{B}_\Omega}(U(f)) \neq 0.$$

If, instead, we choose the sequence $\Phi' = \{\text{sinc}(4\Omega x - n)\}_{n \in \mathbb{Z}}$ for the target space $\mathcal{B}_{2\Omega}$, then $(U, \mathbf{u}(\Psi, \Phi'), \Psi, \Phi')$ with $\mathbf{u}(\Psi, \Phi')$ as in (3.3) has no aliasing error since Φ' is an orthonormal basis for $\mathcal{B}_{2\Omega}$. Furthermore, sampling the input and output functions with arbitrarily higher sampling rate, i.e. representing the functions with respect to the system $\{\text{sinc}(2\bar{\Omega}x - n)\}_{n \in \mathbb{Z}}$ with $\bar{\Omega} > 2\Omega$, yields no aliasing error since $\{\text{sinc}(2\bar{\Omega}x - n)\}_{n \in \mathbb{Z}}$ constitutes a frame for $\mathcal{B}_{2\Omega}$.

E Additional Operator Learning Architectures

DeepONets. Following [18] and for simplicity of exposition, we consider *sensors*, located on a uniform grid on $[-1, 1]$ with grid size $1/(2N + 1)$. We assume that the input function $f \in \mathcal{P}_N$, where \mathcal{P}_N is defined as in SM C with orthonormal basis Ψ_N chosen according to (C.2). Denote the corresponding synthesis and analysis operators as $T_{\Psi_N}, T_{\Psi_N}^*$, respectively. Let $\tau_k : \mathbb{R} \rightarrow \mathbb{R}$, for $1 \leq k \leq K$ be neural networks that form the so-called *trunk nets* in a DeepONet [18] and denote $\mathcal{Q}_K = \text{span}\{\tau_k : k = 1, \dots, K\}$. Clearly $\mathcal{T}_K = \{\tau_k\}_{k=1}^K$ constitutes a *frame* for \mathcal{Q}_K and we can denote the corresponding synthesis and analysis operators by $T_{\mathcal{T}_K}, T_{\mathcal{T}_K}^*$. Then a DeepONet [18] is given by the composition $T_{\mathcal{T}_K} \circ \mathcal{N} \circ T_{\Psi_N}^*$, with $\mathcal{N} : \mathbb{R}^{2N+1} \rightarrow \mathbb{R}^K$, being a neural network of the form (4.3) that is termed as the *branch net* of the DeepONet. Written in this manner, DeepONet satisfies Definition 3.4 as it corresponds to the following commutative diagram,

$$\begin{array}{ccccccc}
 \mathcal{P}_N & \xrightarrow{T_{\Psi_N}^*} & \mathbb{R}^{2N+1} & \xrightarrow{\mathcal{N}} & \mathbb{R}^K & \xrightarrow{T_{\mathcal{T}_K}} & \mathcal{Q}_K \\
 T_{\Psi_N} \uparrow & & \uparrow \text{Id} & & \uparrow \text{Id} & & \uparrow T_{\mathcal{T}_K}^* \\
 \mathbb{R}^{2N+1} & \xrightarrow{\text{Id}} & \mathbb{R}^{2N+1} & \xrightarrow{\mathcal{N}} & \mathbb{R}^K & \xrightarrow{\text{Id}} & \mathbb{R}^K
 \end{array}$$

However, it is essential to emphasize that the choice of the underlying function spaces is essential in regard to the Definition 3.4 of Representation equivalent Neural Operators. For instance, if the sensors are *randomly* distributed rather than located on a uniform grid, then it can induce aliasing errors for DeepONets (see also [14] for a discussion on this issue). Similarly, DeepONets are only ReNOs with respect to the function space \mathcal{Q}_K as the target space. Changing the target function space to another space, say for instance $\mathcal{P}_{N'}$, for some N' , will lead to aliasing errors as the trunk nets do not necessarily form a frame for the space of bandlimited 2-periodic functions.

F Empirical Analysis

F.1 Illustration of the effect of the activation function

We consider the same setting as for the FNO in Section 4 and **SM C**, i.e. considering $f \in \mathcal{P}_K, K = 20$ to be both periodic and bandlimited. On the upmost plot of Figure 5, we observe the values of f , as well as $\text{relu}(f)$ and $\text{gelu}(f)$. In this case, pointwise samples on the grid of size $2K + 1$ are enough to characterize f , as its Fourier coefficients are zero above the Nyquist frequency K . However, as we can clearly observe on the lower plot, this is no longer the case for functions $\text{relu}(f)$ and $\text{gelu}(f)$, and as a consequence the continuous functions are no longer represented uniquely on the grid, and aliasing errors occur.

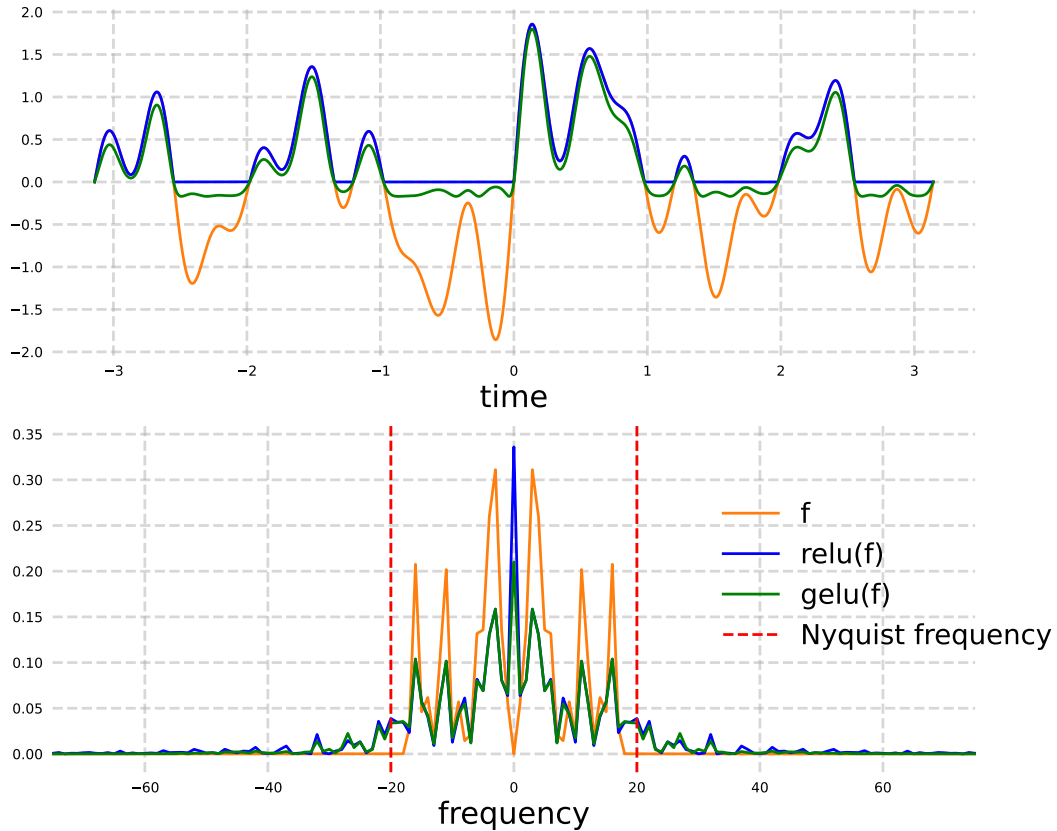


Figure 5: The activation function increases the bandwidth of the input function beyond the Nyquist frequency, causing aliasing errors.

F.2 Details for Experimental Analysis Section 5

Implementation of the experiments is available at

■ <https://github.com/emited/reno.git>

Below we explain how the target operator $Q : \mathcal{H} \rightarrow \mathcal{H}$ is constructed. In practice the support of the function is \mathbb{R}^d , $d = 2$, however for ease of exposition we will present the one-dimensional case $d = 1$, as the multi-dimensional case follows easily.

Here, \mathcal{H} corresponds to the space of bandlimited and periodic functions \mathcal{P}_K , for $K = 30$, as defined in **SM C**. As any function $f \in \mathcal{P}_K$ can be put in correspondence with its pointwise values on the grid $\{f(\frac{k}{2K+1})\}_{k=-K}^K$, see [23, Section 5.5.2], we simply sample these values for the input and output functions from the normal distribution $\mathcal{N}(0, \frac{1}{3})$ with a fixed seed. From this, 128 input-target pairs (f_i, g_i) are produced, and each model is trained to regress on the data. Once trained, the models are tested with respect to different representations of the training data, and the discrete aliasing error map is computed for each of these resolutions as explained in the main paper.

References

- [1] J. J. Benedetto. Irregular sampling and frames. *Wavelets: A Tutorial in Theory and Applications*, 2:445–507, 1992.
- [2] M. Bertero, P. Bocacci, and C. De Mol. *Introduction to inverse problems in imaging*. CRC press, 2021.
- [3] K. Bhattacharya, B. Hosseini, N. B. Kovachki, and A. M. Stuart. Model Reduction And Neural Networks For Parametric PDEs. *The SMAI journal of computational mathematics*, 7:121–157, 2021.
- [4] S. Cao. Choose a transformer: Fourier or galerkin. In *35th conference on neural information processing systems*, 2021.
- [5] T. Chen and H. Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995.
- [6] O. Christensen. *Frames and bases: An introductory course*. Springer Science & Business Media, 2008.
- [7] T. A. Driscoll, N. Hale, and L. N. Trefethen. *Chebfun Guide*. Pafnuty Publishers, Oxford, 2014.
- [8] L. C. Evans. *Partial differential equations*, volume 19. American Mathematical Soc., 2010.
- [9] V. Fanaskov and I. Oseledets. Spectral neural operators. *arXiv preprint arXiv:2205.10573*, 2022.
- [10] V. Isakov. *Inverse Problems for Partial Differential Equations*. Springer, 2017.
- [11] G. Kissas, J. H. Seidman, L. F. Guilhoto, V. M. Preciado, G. J. Pappas, and P. Perdikaris. Learning operators with coupled attention. *Journal of Machine Learning Research*, 23(215):1–63, 2022.
- [12] N. Kovachki, S. Lanthaler, and S. Mishra. On universal approximation and error bounds for fourier neural operators. *Journal of Machine Learning Research*, 22:Art–No, 2021.
- [13] N. B. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. M. Stuart, and A. Anandkumar. Neural operator: Learning maps between function spaces. *CoRR*, abs/2108.08481, 2021.
- [14] S. Lanthaler, S. Mishra, and G. E. Karniadakis. Error estimates for DeepONets: A deep learning framework in infinite dimensions. *Transactions of Mathematics and Its Applications*, 6(1):tnac001, 2022.

- [15] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.
- [16] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. M. Stuart, and A. Anandkumar. Neural operator: Graph kernel network for partial differential equations. *CoRR*, abs/2003.03485, 2020.
- [17] Z. Li, N. B. Kovachki, K. Azizzadenesheli, B. Liu, A. M. Stuart, K. Bhattacharya, and A. Anandkumar. Multipole graph neural operator for parametric partial differential equations. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 6755–6766. Curran Associates, Inc., 2020.
- [18] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.
- [19] M. Prasthofer, T. De Ryck, and S. Mishra. Variable input deep operator networks. *arXiv preprint arXiv:2205.11404*, 2022.
- [20] A. Quarteroni and A. Valli. *Numerical approximation of Partial differential equations*, volume 23. Springer, 1994.
- [21] J. H. Seidman, G. Kissas, P. Perdikaris, and G. J. Pappas. NOMAD: Nonlinear manifold decoders for operator learning. *arXiv preprint arXiv:2206.03551*, 2022.
- [22] M. Unser. Sampling-50 years after Shannon. *Proceedings of the IEEE*, 88(4):569–587, 2000.
- [23] M. Vetterli, J. Kovacevic, and V. Goyal. *Foundations of Signal Processing*. Cambridge University Press, 2014.