REPORT ON THE PAPER "CONVOLUTIONAL NEURAL OPERATOR FOR ROBUST AND ACCURATE LEARNING OF PDES"

DEVANSH TRIPATHI[1]

ETH Zürich

ABSTRACT. In the paper [4], the author argues that the convolution based neural network architectures – believed to be inconsistent in function space – have been largely ingnored in the context of learning solution operators of PDEs. The author present a novel framework termed as convolutional neural operators (CNOs) that is designed specifically to preserve its underlying continuous nature, even when implemented in a discretized form on a computer. The author also proved a universal approximation result for CNOs.

1. Introduction

Given the ubiquitous nature of partial differential equations (PDEs) as mathematical models in the science and engineering, it becomes important to develop method to approximate the solutions to a PDE with less computational cost. There are well-established numerical methods such as finite differences, finite elements, finite volumes and spectral methods that have been successfully used to approximate PDE solution operator. Hwoever, the high computational cost of these methods, particularly in high dimensions and for *many query* problems such Uncertainity Quantification (UQ), inverse problems etc. calls upon the design of *fast, robus and accurate* surrogates.

As *operators* are the objects of interest in solving PDEs, learning such operators from data which is loosely termed as *operator learning*, has emerged as a dominant paradigm in recent years. As it is argued in a recent paper [1], a structure-preserving operator learning algorithm or *representation equivalent neural operator* has to respect some form of continuous-discrete equivalence (CDE) in order to learn the underlying operator, rather than just a discrete representation of it. Failure to respect such a CDE can lead to the so-called aliasing errors [1] and affect model performance at multiple discrete resolutions.

The naive use of convolutional neural networks (CNNs) in the context of operator learning, see [1, 6, 3] on how using CNNs for operator learning leads to results that heavily rely on the underlying grid resolution. The author has made the following contributions in this paper:

- The author proposes novel modifications to CNNs in order to enforce structure-preserving continuous-discrete equivalence (CDE) and enable the genuine, alias-free, learning of operators. The resulting architecture, termed as *Convolutional Neural Operator*(CNO), is provided as novel *operator* adaptation of the widely used U-Net architecture.
- The author has shown that CNO is a *representation equivalent neural operator* in the sense of [1], and also proved a universality result for CNOs to any desired accuracy.
- CNO has been tested on a *novel* set of benchmarks, known as *Representative PDE Benchmarks*(RPB), that span across a variety of PDEs ranging from linear elliptic and hyperbolic to nonlinear parabolic and hyperbolic PDEs, with possibly *multiscale solutions*.

---

[1]Seminar für Angewandte Mathematik, HG E 62.2, Rämistrasse 101, 8092 Zürich, Switzerland
devansh.tripathi@sam.math.ethz.ch.

2. Convolutional Neural Operator

Setting. For simplicity, we will focus here on the two-dimensional case by specifying the underlying domain as $D = \mathbb{T}^2$, being the 2-d torus. Let $\mathcal{X} = H^r(D, \mathbb{R}^{d_{\mathcal{X}}}) \subset \mathcal{Z}$ and $\mathcal{Y} = H^s(D, \mathbb{R}_{d_{\mathcal{Y}}})$ be the underlying function spaces, where $H^{r,s}(D, \cdot)$ are sobolev spaces of order r and s. Without loss of generality, we set $r = s$ hereafter. Our aim would be to aproximate *continuous operators* $\mathcal{G}^\dagger : \mathcal{X} \to \mathcal{Y}$ from data pairs $(u_i, \mathcal{G}^\dagger(u_i))_{i=1}^M \in \mathcal{X} \times \mathcal{Y}$. We furthur assume that there exists a *modulus of continuity* for the operator i.e.,

$$\|\mathcal{G}^\dagger(u) - \mathcal{G}^\dagger(v)\|_{\mathcal{Y}} \leq \omega(\|u - v\|_{\mathcal{Z}}), \qquad \forall u, v \in \mathcal{X}, \tag{2.1}$$

with $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ being a monotonically increasing function with $\lim_{y \to 0} \omega(y) = 0$ (implies that the operator $\mathcal{G}^\dagger$ is uniformly continuous) The underlying operator $\mathcal{G}^\dagger$ can corresponds to solution operators for PDEs but is more general that that and encompasses examples such as those arising in inverse problems, for instance in imaging.

Bandlimited Approximation. As argued in the paper [1] that Sobolev spaces such as H^r are, in a sense, too large to allow for any *continuous-discrete equivalence* (CDE), i.e. equivalence between the underlying operator and its discrete representations, which is necessary for robust operator learning. We have to consider small subspaces of H^r which allow for such CDEs. We choose the space of *bandlimited functions* defined by,

$$B_w(D) = \{f \in L^2(D) : \operatorname{supp}\hat{f} \subseteq [-w, w]^2\}, \tag{2.2}$$

for some $w > 0$ and with $\hat{f}$ denoting the Fourier transform of f. It is shown in appendix A.1 that for any $\varepsilon > 0$, there exists a w, large enough depending on r, and a continuous operator $\mathcal{G}^* : B_w(D) \to B_w(D)$, such that $\|\mathcal{G}^\dagger - \mathcal{G}^*\| \leq \varepsilon$, with $\|\cdot\|$ denoting the corresponding operator norm. Along with that A.2 shows that we can define discrete versions of $\mathcal{G}^*$ using the underlying *sinc* basis for bandlimited functions and establish a continuously-discrete equivalence for it.

Definition of CNO. Given above context, our goal will be approximate the operator $\mathcal{G}^*$ in a *structure-preserving manner* i.e. as the underlying operator maps between spaces of bandlimited functions, we will construct out operator approximation architecture to also map bandlimited functions to bandlimited functions, thus respecting the continuous-discrete equivalence.

We denote the operator $\mathcal{G} : B_w(D) \to B_w(D)$ as a *convolutional neural operator* (CNO) which we define as a compositional mapping between functions as

$$\mathcal{G} : u \mapsto P(u) = v_0 \mapsto v_1 \mapsto \ldots v_L \mapsto Q(v_L) = \overline{u}, \tag{2.3}$$

where

$$v_{l+1} = \mathcal{P}_l \circ \Sigma_l \circ \mathcal{K}_l(v_l), \qquad 1 \leq l \leq L - 1. \tag{2.4}$$

From 2.3, we see that first, the input function $u \in B_w(D)$ is lifted to the latent space of bandlimited functions through a *lifting layer*:

$$P : \left\{u \in B_w(D, \mathbb{R}^{d_{\mathcal{X}}})\right\} \to \left\{u \in B_w(D, \mathbb{R}^{d_0})\right\}$$

Here, $d_0 > d_{\mathcal{X}}$ is the number of channels in the lifted, latent space. The lifting operation is performed by a convolution operator which will be defined below.

Then, the lifted function is processed through the composition of a series of mappings between functions (layers), with each layer consisting of three elementary mappings, i.e. $\mathcal{P}_l$ is either the *upsampling* or *downsampling* operator, $\mathcal{K}_l$ is the convolution operator and Σ_l is the activation operator Q, defined as

$$Q : \left\{v_L \in B_w(D, \mathbb{R}^{d_L})\right\} \to \left\{\overline{u} \in B_w(D, \mathbb{R}^{d_{\mathcal{Y}}})\right\}.$$

The projection operation is also performed by a convolution operator defined below.

Convolution Operator. For simplicity, *single channel* version of the convolution operator K_l is presented here. See appendix A.3 for *multi-scale* version. Convolution operations are performed with discrete kernals

$$K_w = \sum_{i,j=1}^{k} k_{ij} \cdot \delta_{z_{ij}}$$

defined on a $s \times s$ uniform grid on D with grid size $\geq 1/2w$, in-order to satisfy the requirements of the Whittaker-Shannon Kotelnikov sampling theorem [5], and z_{ij} being the resulting grid points, $k \in \mathbb{N}$ being kernel size and δ_x denoting the Dirac measure at point $x \in D$. The convolution operator for a *single-channel* $\mathcal{K}_w : B_w(D) \to B_w(D)$ is defined by

$$\mathcal{K}_w f(x) = (K_w \star f)(x) = \int_D K_w(x-y)f(y)dy = \sum_{i,j=1}^{k} k_{i,j} f(x - z_{ij}), \qquad \forall x \in D,$$

where the last identity arises from the fact that $f \in B_w$. Thus, our convolution operator is directly parameterized in physical space, in contrast to the Fourier space parameterization of convolution in the FNO architecture of [3]. Hence, our parameterization is of a *local* nature.

Upsampling and Downsampling Operator. For some $\overline{w} > w$, we can *upsample* a function $f \in B_w$ to the *higher band* $B_{\overline{w}}$ by simple setting,

$$\mathcal{U}_{w,\overline{w}} : B_w(D) \to B_{\overline{w}}(D), \qquad \mathcal{U}_{w,\overline{w}} f(x) = f(x), \qquad \forall x \in D.$$

On the other hand, for some $\underline{w} < w$, we can *downsample* a function $f \in B_w$ to the *lower band* $B_{\underline{w}}$ by setting $\mathcal{D}_{w,\underline{w}} : B_w(D) \to B_{\underline{w}}(D)$, defined by

$$\mathcal{D}_{w,\underline{w}} f(x) = \left(\frac{\underline{w}}{w}\right)^2 (h_{\underline{w}} \star f)(x) = \left(\frac{\underline{w}}{w}\right)^2 \int_D h_{\underline{w}}(x-y)f(y)dy, \qquad \forall x \in D,$$

where $\star$ is the convolution operation on functions defined above and $h_{\underline{w}}$ is the so-called *interpolation sinc filter*:

$$h_w(x_0, x_1) = \text{sinc}(2wx_0) \cdot \text{sinc}(2wx_1), \qquad (x_0, x_1) \in \mathbb{R}^2. \tag{2.5}$$

Activation Layer. Naturally, one can apply the activation function pointwise to any function. However, it is well-known that such an application will no longer respect the band-limits of the underlying function space and generate *aliasing error* [1]. In particular, nonlinear activations can generate features at arbitrarily high frequencies. Our aim is to respect the underlying CDE, we will modulate the application of the activation function so that the resulting outputs falls within desired band limits.

First, we will upsample the input function $f \in B_w$ to a higher bandlimit $\overline{w} > w$, then apply the activation function pointwise and finally downsample the result back to the original bandlimit w. Implicitly assuming that $\overline{w}$ is large enough such that $\sigma(B_w) \subset B_{\overline{w}}$, we define the activation layer as,

$$\Sigma_{w,\overline{w}} : B_w(D) \to B_w(D), \qquad \Sigma_{w,\overline{w}} f(x) = \mathcal{D}_{\overline{w},w}(\sigma \circ \mathcal{U}_{w,\tilde{w}} f)(x), \qquad \forall x \in D. \tag{2.6}$$

Instantiation through an Operator U-Net architecture. The above blocks are assembled together in the form of an Operator U-Net architecture that has bandlimited functions as input and outputs. We also need other ingredients such as *skip connections* through *ResNet* blocks of the form, $\mathcal{R}_{w,\overline{w}} : B_w(D, \mathbb{R}^d) \to B_w(D, \mathbb{R}^d)$, such that

$$\mathcal{R}_{w,\overline{w}}(v) = v + \mathcal{K}_w \circ \Sigma_{w,\overline{w}} \circ \mathcal{K}_w(v), \qquad \forall v \in B_w(D, \mathbb{R}^d). \tag{2.7}$$

we also need the *Invariant blocks* of the form , $\mathcal{I}_{w,\overline{w}} : B_w(D, \mathbb{R}^d) \to B_w(D, \mathbb{R}^d)$ such that

$$\mathcal{I}_{w,\overline{w}}(v) = \Sigma_{w,\overline{w}} \circ \mathcal{K}_w(v), \qquad \forall v \in B_w(D, \mathbb{R}^d). \tag{2.8}$$

Continuous-Discrete Equivalence for CNO. We have defined CNO as an operator that maps bandlimited functions to bandlimited functions. In practice, CNO has to be implemented in a discrete manner with *discretized versions* of each of the above-defined elementary operations. We prove the following proposition:

Proposition 1. *Convolutional Neural Operator* $\mathcal{G} : B_w(D, \mathbb{R}^{d_{\mathcal{X}}}) \to B_w(D, \mathbb{R}^{d_{\mathcal{Y}}})$ *is a Representation equivalent neural operator or ReNO, in the sense of Defintion 6.*

Proof. TODO ■

3. Universal Approximation by CNOs

We consider the following abstract PDE in the domain $D = \mathbb{T}^2$,

$$\mathcal{L}(u) = 0, \qquad \mathcal{B}(u) = 0, \tag{3.1}$$

with $\mathcal{L}$ being the differential operator and $\mathcal{B}$ a boundary operator. We assume that the differential operator $\mathcal{L}$ only depends on the coordinate x through a *coefficient* function $a \in H^r(D)$. The corresponding *solution* operator is denoted by $\mathcal{G}^{\dagger} : \mathcal{X}^* \subset H^r(D) \to H^r(D) : a \mapsto u$, with u begin the solution of the PDE (3.1). We assume that $\mathcal{G}^{\dagger}$ is continuous. Moreover, we also assume the following modulus of continuity,

$$\left\| \mathcal{G}^{\dagger}(a) - \mathcal{G}^{\dagger}(a') \right\|_{L^p(\mathbb{T}^2)} \leq \omega \left(\|a - a'\|_{H^{\sigma}(\mathbb{T}^2)} \right), \tag{3.2}$$

for some $p \in \{2, \infty\}$ and $0 \leq \sigma \leq r - 1$, and where $\omega : [0\infty) \to [0, \infty)$ is a monotonously increasing function with $\lim_{y \to 0} \omega(y) = 0$.

Appendix A. **Detailed proofs**

A.1. **Approximation of Operators mapping between Sobolev spaces by operators mapping between spaces of bandlimited functions.** We prove that one can approximate any continuous operator $\mathcal{G}^{\dagger} : \mathcal{X} \to \mathcal{Y}$ by an operator mapping between spaces of bandlimited functions to arbitrary accuracy. We obtain this result by dsicarding the high-frequency components, higher than the frequency w, of both the input and the output of $\mathcal{G}^{\dagger}$ by a Fourier projection P_w. For orthogonal Fourier projections and also trigonometric polynomial interpolation, the following result on the accuracy of the projection holds,

Lemma (A.1). *Given* $\sigma, r \in \mathbb{N}_0$ *with* $r > d/2$ *and* $r \geq \sigma$, *and* $f \in C^r(\mathbb{T}^d)$ *it holds for every* $w \in \mathbb{N}$ *that,*

$$\|f - P_w(f)\|_{H^{\sigma}(\mathbb{T}^d)} \leq C(r, d) w^{-(r-\sigma)} \|f\|_{H^r(\mathbb{T}^d)}, \tag{A.1}$$

for a constant $C(r, d) > 0$ *that only depends on* r *and* d.

By choosing an appropriate frequency cutoff and then discarding the high frequencies of the input and output of $\mathcal{G}^{\dagger}$ one can approximate $\mathcal{G}^{\dagger}$ to arbitrary accuracy as shown in the result below.

Lemma (A.2). *For any* $\epsilon, B > 0$ *there exist* $w \in \mathbb{N}$ *such that* $\|\mathcal{G}^{\dagger}(a) - P_w\mathcal{G}^{\dagger}(P_wa)\|_{L^2(D)} \leq \epsilon$ *for all* $a \in H^r(D)$ *with* $\|a\|_{H^r(D)} \leq B$.

Proof. Using Lemma A.1 and stability of $\mathcal{G}^{\dagger}$, equation 2.1, we have

$$\|\mathcal{G}^{\dagger}(a) - P_w\mathcal{G}^{\dagger}(P_wa)\|_{L^2} \leq \|\mathcal{G}^{\dagger}(a) - P_w\mathcal{G}^{\dagger}(a)\|_{L^2} + \|P_w\mathcal{G}^{\dagger}(a) - P_w\mathcal{G}^{\dagger}(P_wa)\|_{L^2}$$

Since, $\mathcal{G}^{\dagger} : H^r(D; \mathbb{R}^{d_{\mathcal{X}}}) \to H^s(D; \mathbb{R}^{d_{\mathcal{Y}}})$, we have $\mathcal{G}(a) \in H^s(D; \mathbb{R}^{d_{\mathcal{Y}}})$. In order to apply Lemma A.1, we need to show that $\mathcal{G}^{\dagger}(a) \in C^r(D)$. Note that $H^s(D) \hookrightarrow C^r(D)$ for all $s > r + d/2$ which

will imply that $\mathcal{G}^\dagger(a) \in C^r(D)$. Also, note that P_w(Fourier projection operator) is *non-expansive* (L^2-norm is less than 1). Taking $\sigma = 0$ in Lemma A.1, we have

$$
\begin{aligned}
\left\|\mathcal{G}^\dagger(a) - P_w\mathcal{G}^\dagger(P_wa)\right\|_{L^2} &\leq \left\|\mathcal{G}^\dagger(a) - P_w\mathcal{G}^\dagger(a)\right\|_{L^2} + \left\|P_w\mathcal{G}^\dagger(a) - P_w\mathcal{G}^\dagger(P_wa)\right\|_{L^2} \\
(\because H^0(D) = L^2(D)) \qquad &= \left\|\mathcal{G}^\dagger(a) - P_w\mathcal{G}^\dagger(a)\right\|_{H^0} + \left\|P_w\mathcal{G}^\dagger(a) - P_w\mathcal{G}^\dagger(P_wa)\right\|_{L^2} \\
\text{(Lemma A.1 and non-expansive)} \qquad &\lesssim w^{-r}\left\|\mathcal{G}^\dagger(a)\right\|_{H^r} + \left\|\mathcal{G}^\dagger(a) - \mathcal{G}^\dagger(P_wa)\right\|_{L^2} \\
&\lesssim w^{-r}\|\mathcal{G}^\dagger\|_{op}\|a\|_{H^r} + \omega(\|a - P_wa\|_{H^\sigma}) \\
&\lesssim w^{-r}\|\mathcal{G}^\dagger\|_{op}\|a\|_{H^r} + \omega(Cw^{-(r-\sigma)}\|a\|_{H^r}).
\end{aligned}
$$

(A.2)

∎

Since above relation is true for all a such that $\|a\|_{H^r} \leq B$, it follows that for large enough w:

$$
\sup_{\|a\|_{H^r} \leq B} \left\|\mathcal{G}^\dagger(a) - P_w\mathcal{G}^\dagger(P_wa)\right\|_{L^2} \leq \epsilon.
$$

(A.3)

Given that both $P_wa \in B_w(D)$ and $P_w\mathcal{G}^\dagger(P_wa) \leq B_w(D)$, a consequence of the above lemma is the existence of an operator $\mathcal{G}^*"B_w(D) \to B_w(D) : a \mapsto P_w(\mathcal{G}^\dagger(a))$ that can approximate $\mathcal{G}^\dagger$ arbitrarily well. Hence $\left\|\mathcal{G}^\dagger - \mathcal{G}^*\right\|_{op} \leq \epsilon$, where the operator are considered as mapping from and to $B_w(D) \cap H^r(D)$ equipped with the $H^r(D)$-norm.

A.2. **Continuous-Discrete Equivalence for Operator $\mathcal{G}^*$.** For every $w > 0$, we denote by $B_w(\mathbb{R}^2)$ the space of multivariate bandlimited functions

$$
B_w(\mathbb{R}^2) = \{f \in L^2(\mathbb{R}^2) : \mathrm{supp}\hat{f} \subset eq[-w, w]^2\},
$$

where $\hat{f}$ denotes the Fourier transform on $L^1(\mathbb{R})$

$$
\hat{f}(\xi) := \int_{\mathbb{R}} f(x)e^{-2\pi ix\xi}dx, \qquad \xi \in \mathbb{R},
$$

which extends to $L^2(\mathbb{R})$ by a classical density argument ($L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n), n \geq 1$.) The set $\Psi_w = \{\mathrm{sinc}(2wx_1 - m) \cdot \mathrm{sinc}(2wx_2 - n)\}_{m,n\in\mathbb{Z}}$ constitutes an orthonormal basis for $B_w(\mathbb{R}^2)$. The bounded operator

$$
T_{\Psi_w} : l^2(\mathbb{Z}^2) \to B_w(\mathbb{R}^2), \ T_{\Psi_w}(c_{m,n}) = \sum_{m,n\in\mathbb{Z}} c_{m,n}\mathrm{sinc}(2w \cdot -m) \cdot \mathrm{sinc}(2w \cdot -n),
$$

which reconstructs a function from its basis coefficients, is called *synthesis operator,* and its adjoint

$$
T_{\Psi_w}^* : B_w(\mathbb{R}^2) \to l^2(\mathbb{Z}^2), \qquad T_{\Psi_w}^*f = \left\{f\left(\frac{m}{2w}, \frac{n}{2w}\right)\right\}_{m,n\in\mathbb{Z}},
$$

which extracts basis coefficients from an underlying function, is called *analysis operator*. Every bandlimited function can be uniquely and stably recovered from its sampled values $\left\{f\left(\frac{m}{2w}, \frac{n}{2w}\right)\right\}$ for $m, n \in \mathbb{Z}$ via the reconstruction formula

$$
f(x_1, x_2) = T_{\Psi_w}T_{\Psi_w}^*f(x_1, x_2) = \sum_{m,n\in\mathbb{Z}} f\left(\frac{m}{2w}, \frac{n}{2w}\right)\mathrm{sinc}(2wx_1 - m) \cdot \mathrm{sinc}(2wx_2 - n), \quad \text{(A.4)}
$$

and we say that there is a *continuous-discrete equivalence (CDE)* between f and its samples $\left\{f\left(\frac{m}{2w}, \frac{n}{2w}\right)\right\}$. In general, every bandlimited function $f \in B_w(\mathbb{R}^2)$ cam be uniquely and stably recovered from its sample values $\{f(mT, nT)\}_{m,n\in\mathbb{Z}}$ if the *sampling rate* or reciprocal of grid size, $1/T$ is greater or equal than the *Nyquist rate* $2w$. This simply follows from the fact that $B_w(\mathbb{R}^2) \subset B_{w'}(\mathbb{R}^2)$ for every $w' > w$. On the contrary, reconstructing $f \in B_w$ at a sampling rate below the Nyquist rate, i.e. $1/T < 2w$, results in a non-zero value for the *aliasing error function:*

$$
\epsilon(f) = f - T_{\Psi_{\frac{1}{2T}}} T_{\Psi_{\frac{1}{2T}}}^* f,
$$

and the associated *aliasing error* $\|\epsilon\|_2$ (see defintion below).

Definition 1 (Aliasing for bandlimited functions)**.** 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

$$\varepsilon(f) = f - \mathcal{P}_{B_\Omega} f, \qquad \|\varepsilon(f)\|_2 = \|f - \mathcal{P}_{B_\Omega} f\|_2.$$

where $\mathcal{P}_{B_\Omega} : L^2(\mathbb{R}) \to B_\Omega$ is the *orthogonal projection operator* onto B_Ω given by $P_{B_\Omega} f = \sum_{n\in\mathbb{Z}} \langle f, \phi_n \rangle \phi_n$. If the aliasing error $\varepsilon(f)$ is zero, i.e. if $f \in B_\Omega$, we say that there is a continuous-discrete equivalence (CDE) between f and its samples $\{f(n/2\Omega)\}_{n\in\mathbb{Z}}$.

Let $\mathcal{G}^*$ be a (possibly) non-linear operator between band-limited spaces, i.e. $\mathcal{G}^* : B_w(\mathbb{R}^2) \to B_{w'}(\mathbb{R}^2)$, for some $w, w' > 0$. As given in C, the concepts of continuous-discrete equivalence (CDE) and aliasing error can be adapted to the operator $\mathcal{G}^*$. The continuous operator $\mathcal{G}^*$ is uniquely determined by a map $\mathfrak{g}_{\Psi_w, \Psi_{w'}} : \ell^2(\mathbb{Z}^2) \to \ell^2(\mathbb{Z}^2)$ if the aliasing error operator.

$$\varepsilon = \mathcal{G}^* - T_{\Psi_{w'}} \circ \mathfrak{g}_{\Psi_w, \Psi_{w'}} \circ T^*_{\Psi_w} \tag{A.5}$$

is identically zero, and we say that $\mathcal{G}^*$ and $\mathfrak{g}_{\Psi_w, \Psi_{w'}}$, satisfy a continuous-discrete equivalence (Definition 4). Equivalently, the diagram

FIGURE 1

commutes, i.e. the black and the blue directed paths in the diagram lead to the same result. In the latter case, since $T^*_{\Psi_w} \circ T_{\Psi_w}$ is identity operator from $\ell^2(\mathbb{Z}^2)$ onto itself ($\because \Psi_w$ is taken to be orthonormal basis set, otherwise it is Gram matrix $G_{mn} := \langle \Psi_n, \Psi_m \rangle$), equation A.5 forces the discretization $\mathfrak{g}_{\Psi_w, \Psi_{w'}}$, to be defined as

$$\mathfrak{g}_{\Psi_w, \Psi_{w'}} := T^*_{\Psi_{w'}} \circ \mathcal{G}^* \circ T_{\Psi_w}. \tag{A.6}$$

The above defintion is motivated from the fact that Ψ_w is orthonormal set and the equation A.5 when $\varepsilon = 0$.

Remark 1. Note that in general T_{Ψ_w} is not invertible. When the frame sequence is orthonormal, it is invertible and its inverse is same as its adjoint.

Remark 2. While isolating $\mathfrak{g}_{\Psi_w, \Psi_{w'}}$ on LHS in equation A.5, the term $T^{-1}_{\Psi_{w'}}$ will appear and it has been replaced to $T^*_{\Psi_{w'}}$ in equation A.6 because of remark 1. Why? That is how we define it.

The definition A.6 is equivalent to say that the diagram, commutes. In other words, once we fix

FIGURE 2

the discrete representation assocciated to the input and output functions, there exists a unique way to defined a discretization $\mathfrak{g}_{\Psi_w, \Psi_{w'}}$ that is consistent with the continuous operator $\mathcal{G}^*$ and this is given by A.6.

In practice, we may have access to different discrete representations of the input and output functions, e.g. point samples evaluated on different grids, which in the theory amounts to change of reference systems in the functions spaces. For instance, sampling a function $f \in B_w$

on a finer gris $\left\{\left(\frac{m}{2\overline{w}}, \frac{n}{2\overline{w}}\right)\right\}_{m,n\in\mathbb{Z}}, \overline{w} > w$, amounts to representing the function f with respect to the system $\Psi_{\overline{w}} = \{\text{sinc}(2\overline{w}x_1 - m) \cdot \text{sinc}(2\overline{w}x_2 - n)\}_{m,n\in\mathbb{Z}}$, which constitutes an orthonormal basis for $B_{\overline{w}} \supset B_w$. Then, one can define the associated CDE discretization $\mathfrak{g}_{\Psi_{\overline{w}}, \Psi_{\overline{w}'}}$ as in A.6, and by equation A.5, one readily obtains the change of basis formula

$$\mathfrak{g}_{\Psi_{\overline{w}}, \Psi_{\overline{w}'}} = T^*_{\Psi_{\overline{w}'}} \circ T_{\Psi_{w'}} \circ \mathfrak{g}_{\Psi_w, \Psi_{w'}} \circ T^*_{\Psi_w} \circ T_{\Psi_{\overline{w}}}, \tag{A.7}$$

Finally, all the above concepts generalize to every pair of frame sequences (Ψ, Ψ) that span respectively the input and output function spaces, and we refer [1] for a complete exposition.

A.3. Multi-channel versions of elementary operators for CNO.

This section contains the *multi-channel* versions of the elementary mappings which define CNO 2.

Convolution Operator. In the multi-channel settings, discrete kernels K_w are defined on the $d_{in} \times d_{out} \times s^2$ uniform grids on D, where d_{in} is the number of input channels and d_{out} is the number of output channels. Formally, the kernels are defined as

$$K_{w,cl} = \sum_{i,j=1}^{k} k_{i,j,cl} \cdot \delta_{z_{ij}}.$$

where c is the channel index in the input space, while l is the channel index in the output space. Each pair of channels defines corresponding single-channel convolution operation $\mathcal{K}_{w,cl} : B_w(D) \to B_w(D)$. For $a \in B_w(D, \mathbb{R}^{d_{in}})$, the multi-channel convolution operation K_w is defined as

$$(K_w a(x))_l = \sum_{c=1}^{d_{in}} \mathcal{K}_{w,cl} a_c(x), \qquad l = 1, \ldots, d_{out}.$$

Upsampling and Downsampling Operators. To upsample a signal $a \in B_w(D, \mathbb{R}^d)$ with d channels from the bandlimit $w > 0$ to the bandlimit $\overline{w} > w$, one should apply the single-channel upsampling operator $\mathcal{U}_{w,\overline{w}}$ to each individual channel of the input signal, independently. Formally, for $a \in B_w(D, \mathbb{R}^d)$, the multi channel upsampling $\mathcal{U}_{w,\overline{w}} : B_w(d, \mathbb{R}^d) \to B_{\overline{w}}(D, \mathbb{R}^d)$ is defined as

$$(\mathcal{U}_{u,\overline{w}} a(x))_c = a_c(x), \qquad \forall x \in D, \quad c = 1, \ldots, d.$$

The downsampling operator of a signal $a \in B_w(D, \mathbb{R}^d)$ from the bandlimit $w > 0$ to the bandlimit $\underline{w} < w$ is defined in a similar manner (independent applications of the single-channel downsampling operators).

Activation layer. The multi-channel version of the activation layer, namely $\Sigma_{w,\overline{w}} : B_w(D, \mathbb{R}^d) \to B_w(D, \mathbb{R}^d)$, is realized by applying the single-channel activation layer to each of the d channels, independently.

A.4. Discrete operators for CNO.

In this section, we define the *discrete versions* of the elementary mappings in 2. Given a *discrete*, multi-channel signal $a_s \in \mathbb{R}^{s \times s \times d}$ on $s \times s \times d$ uniform grid, we will use the notation $a_s[i, j, c]$ to refer to the (i, j)-th coordinate of the c-th channel of the signal, where $i, j = 1, \ldots, s$ and $c = 1, \ldots, d$.

Convolution operator. Assume that instead of a continuous, single-channel signal $a \in B_w(D)$, one has an access only to its sampled version $a_s \in \mathbb{R}^{s \times s}$ on $s \times s$ uniform grid on D. Assume that a_s is to be convolved with a *discrete* kernel $K_w \in \mathbb{R}^{k \times k}$ with $k = 2\hat{k} + 1$. Let $\hat{a}_s \in \mathbb{R}^{(s+2\hat{k}) \times (s+2\hat{k})}$ be an extended version of a_s obtained by circular-padding (domain repeats itself so that kernel can slide through without issues at boundary) to zero-padding (adding 0's) of a_s. The discrete, single-channel convolution $\mathcal{K}_s : \mathbb{R}^{s \times s} \to \mathbb{R}^{s \times s}$ of the signal a_s and the kernel K_w is given by

$$\mathcal{K}_s(a_s) = (a_s \star K_w)[i, j] = \sum_{m,n=-\hat{k}}^{\hat{k}} K_w[m, n] \cdot \hat{a}_s[i - m, j - n], \qquad i, j = 1, \ldots, s,$$

where indices of $\hat{a}$ outside the range $1, \ldots, s$ corresponds to the padded samples. By performing the convolution this way as described above, we make sure that the input and output signals have the same spatial dimension $s \times s$.

Let $a_s \in \mathbb{R}^{s \times s \times d_{in}}$ be a discrete, multi-channel and $K_w \in \mathbb{R}^{k \times k \times d_{in} \times d_{out}}$ a discrete kernel with $k = 2\hat{k} + 1$. The multi-channel convolution of a_s and K_w is defined by

$$(a_s \star K_w)[i, j, l] = \sum_{m,n=-\hat{k}}^{\hat{k}} \sum_{c=1}^{d_{in}} K_w[m, n, c, l] \cdot \hat{a}_s[i - m, j - n, c], \qquad i, j = 1, \ldots, s,$$

where l corresponds to the index of the output channel and c to the index of the input channel.

Upsampling and Downsampling Operators. Here, we will define the discrete upsampling and downsampling operators. For $w > 0$, let h_w be the interpolation *sinc* filter defined in 2.5. For a discrete, single-channel signal $a_s \in \mathbb{R}^{s \times s}$, let $(\tilde{a}_s[n])_{n \in \mathbb{Z}}$ be its periodic extension into infinite length given by $\tilde{a}_s[n] = a_s[n \mod s]$ for $n \in \mathbb{Z}$. The discrete upsampling $\mathcal{U}_{s,N} : \mathbb{R}^{s \times s} \to \mathbb{R}^{Ns \times Ns}$ by an *integer factor* $N \in \mathbb{N}$ of the signal $a_s \in \mathbb{R}^{s \times s}$ is done in *two* phases:

(1) Firstly, in order to increase the number of samples of the signal a_s from s^2 to $(Ns)^2$. One transform the signal a_s into the signal $a_{s,\uparrow Ns}$ obtained by separating each two signal samples of a_s with $N - 1$ zero-valued samples. It holds that $a_{s,\uparrow Ns} \in R^{Ns \times Ns}$ and

$$a_{s,\uparrow Ns}[i, j] = \mathbb{1}_S(i) \cdot \mathbb{1}_S(j) \cdot a_s[i \mod s, j \mod s], \quad i, j = 1, \ldots Ns,$$

where $S = \{1, s + 1, \ldots, (N - 1)s + 1\}$ and $\mathbb{1}_S$ is the indicator function. The matrix representation for $N = 2$ and $s = 2$ will look like this:

$$\begin{bmatrix} \times & 0 & \times & 0 \\ 0 & 0 & 0 & 0 \\ \times & 0 & \times & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where $\times = a_s[i \mod s, j \mod s]$ represent non-zero sampled values. It can be seen that each two signal samples are separated by 0.

(2) Second step is to convolve the periodic extension of $a_{s,\uparrow Ns}$ with the $h_{s/2}$ interpolation filter to eliminate high frequency components. The upsampled signal is formally obtained by

$$\mathcal{U}_{s,N}(a_s)[i, j] = \sum_{n,m \in \mathbb{Z}} \tilde{a}_{s,\uparrow Ns}[n, m] \cdot h_{s/2}(is - ns, js - ms), \quad i, j = 1, \ldots, Ns.$$

The discrete downsampling $\mathcal{D}_{s,N} : \mathbb{R}^{s \times s} \to \mathbb{R}^{s/N \times s/N}$ by an *integer factor* $N \in \mathbb{N}$ of the signal $a_s \in \mathbb{R}^{s \times s}$ is also done in *two* phases (under the assumption that $s/N \in \mathbb{N}$):

(1) First step is to convolve the periodic extension of a_s with the $h_{s/(2N)}$ interpolation filter to eliminate high frequency content. Formally, the first step is defined by

$$a_{s,s/N}[i, j] = \sum_{n,m \in \mathbb{Z}} \tilde{a}_s[n, m] \cdot h_{s/(2N)}(is - ns, js - ms), \quad i, j = 1, \ldots, s/N.$$

(2) Second step is to decrease the sampling rate of $a_{s,N/s}$ by keeping every N-th sample of the signal. The downsampled signal is formally defined by

$$\mathcal{D}_{s,N}(a_s)[i, j] = a_{s,s/N}[(i - 1)s + 1, (j - 1)s + 1], \quad i, j = 1, \ldots, s/N.$$

Multi-channel discrete upsampling and downsampling are performed by independent applications of the corresponding single-channel operator.

Activation layer. Given the definition of the discrete operators, the discrete, single-channel activation layer is defined as

$$\Sigma_s : \mathbb{R}^{s \times s} \to \mathbb{R}^{s \times s}, \quad \Sigma_s(a_s) = \mathcal{D}_{s,N} \circ \sigma \circ \mathcal{U}_{s,N}(a_s),$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is an activation function applied point-wise and $N \in \mathbb{N}$ is a fixed constant. The multi-channel activation layer is performed by independent applications of the single-channel activation layer.

APPENDIX B. **An Introduction to Frame Theory**

Let $\mathcal{H}$ be a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$.

Definition 2 (Frame). A countable sequence of vectors $\{f_i\}_{i \in I}$ in $\mathcal{H}$ is a *frame* for $\mathcal{H}$ if there exists constants $A, B > 0$ such that for all $f \in \mathcal{H}$

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

We say that $\{f_i\}_{i \in I}$ is a *tight frame* if $A = B$ and, in particular, a *Parseval fram* if $A = B = 1$.

Remark 3. The condition in the above definition ensure that f can be stably reconstructed from its inner products with the frame elements. Stable reconstruction here implies the well-posedness and bounded in both direction (no explosion or vanishing of norm).

Remark 4. A frame is complete in $\mathcal{H}$, stable spanning set of $\mathcal{H}$.

Clearly by the Parseval identity, an orthonormal basis for $\mathcal{H}$ is a Parseval frame ($\because \|f\|^2 = \sum_{n=1}^{\infty} |\langle x, x_n \rangle|^2$, by Parseval's identity for a separable Hilbert space, the so called generalization of the Pythagorean theorem). The lower inequality implies that

$$\langle f, f_i \rangle = 0, \forall i \in I \implies f = 0,$$

which is equivalent to

$$\overline{\text{span}}\{f_i : i \in I\} = \mathcal{H}.$$

On the other hand the upper inequality implies that the operator

$$T : l^2(I) \to \mathcal{H}, \qquad T(\{c_i\}_{i \in I}) = \sum_{i \in I} c_i f_i,$$

is bounded with $\|T\| \leq \sqrt{B}$ ([2] Theorem 3.1.3), and we call T the *synthesis operator*. Its adjoint is given by

$$T^* : \mathcal{H} \to \mathcal{H}, \qquad Sf = TT^*f = \sum_{i \in I} \langle f, f_i \rangle f_i,$$

which is bounded, invertible, self-adjoint and positive operator ([2], Lemma 5.1.6). We note that the frame operator is invertible since it is bounded, being a composition of two bounded operator, and the frame property implies that $\|\text{Id} - B^{-1}S\| < 1$, where Id denotes the identity operator. Furthermore, the pseudo-inverse of the synthesis operator is given by

$$T^\dagger : \mathcal{H} \to l^2(I), \qquad T^\dagger f = (\langle f, S^{-1}f_i \rangle)_{i \in I},$$

([2], Theorem 5.3.7) and $\|T^\dagger\| \leq 1/\sqrt{A}$([2], Proposition 5.3.8). The composition $TT^\dagger$ gives the identity operator on $\mathcal{H}$, and consequently every element in $\mathcal{H}$ can be reconstructed via the reconstruction formula

$$f = TT^\dagger f = \sum_{i \in f} \langle f, S^{-1}f_i \rangle f_i = \sum_{i \in I} \langle f, f_i \rangle S^{-1} f_i, \tag{B.1}$$

where the series converge unconditionally. Formula B.1 is known as the *frame decomposition thoerem* ([2], Theorem 5.1.7). In particular, if $\{f_i\}_{i \in I}$ is a tight frame, then $S = A\text{Id}$ and formula B.1 simply reads

$$f = \frac{1}{A} \sum_{i \in I} \langle f, f_i \rangle f_i.$$

On the other hand, the composition $T^\dagger T$ gives the orthogonal projection of $l^2(I)$ onto $\text{Ran}(T^\dagger)$([2], Lemma 2.5.2).

In what follows, we consider sequences which are not complete in $\mathcal{H}$, and consequently are not frames for $\mathcal{H}$, but they are frames for their closed linear span.

Definition 3 (Frame sequence). Let $\{v_i\}_{i \in I}$ be a countable sequence of vectors in $\mathcal{H}$. We say that $\{v_i\}_{i \in I}$ is a frame sequence *if it is a frame for* $\overline{span}\{v_i : i \in I\}$.

Remark 5. A frame sequence may not span all of $\mathcal{H}$, but it is a frame for the subspace of $\mathcal{H}$ which it does span.

A frame sequence $\{v_i\}_{i \in I}$ in $\mathcal{H}$ with synthesis operator $T : l^2(I) \to \overline{span}\{v_i : i \in I\}$ is a frame for $\mathcal{H}$ if and only if T^* is injective, whilst in general T^* is not surjective and consequently T is not injective. We denote $\mathcal{V} = \overline{span}\{v_i : i \in I\}$. Then, the orthogonal projection of $\mathcal{H}$ onto $\mathcal{V}$ is given by

$$\mathcal{P}_\mathcal{V} f = T T^\dagger = \sum_{i \in I} \langle f, S^{-1} v_i \rangle v_i,$$

where $S : \mathcal{V} \to \mathcal{V}$ denotes the frame operator. Hence, reconstruction formula B.1 holds if and only if $f \in \mathcal{V}$.

B.1. **Condition** $u : \mathbf{Ran} \, T_\Psi^\dagger \to \mathbf{Ran} \, T_\Phi^\dagger$**.** Once the discretization is choosen in the input and output spaces, which amounts to choosing a frame pair $(\Psi \subseteq \mathcal{H}, \Phi \subseteq \mathcal{K})$, we want to define a mapping $u : \ell^2(I) \to \ell^2(K)$ which handles such discrete representations. By frame decomposition theorem B.1, every function in $\mathcal{H}$ is uniquely determined by a sequence in $\mathrm{Ran} \, T_\Phi^\dagger$, and analogously every function in $\mathcal{K}$ is uniquely determined by a sequence in $\mathrm{Ran} \, T_\Phi^\dagger$. Hence, it is sufficient to define a mapping from $\mathrm{Ran} \, T_\Psi^\dagger$ into $\mathrm{Ran} \, T_\Phi^\dagger$. We ensure that when two different discretization both yield zero aliasing, the change of frame formula holds, and thus the representation equivalence error is zero.

Appendix C. **Alias-Free Framework for Operator Learning**

In this section, we will extend this concept of aliasing of function to operators. Let U be a operator mapping between infinite-dimensional function spaces and u be a discrete representation mapping.

Setting. Let $U : \mathrm{Dom} \, U \subseteq \mathcal{H} \to \mathcal{K}$ be an operator between two separable Hilbert spaces, and let $\Psi = \{\psi_i\}_{i \in I}$ and $\Psi = \{\phi_k\}_{k \in K}$ 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 $\mathcal{M}_\Psi := \overline{span}\{\psi_i : i \in I\}$ and $\mathcal{M}_\Phi := \overline{span}\{\phi_i : k \in K\}$. We note that by classical frame theory [2], the pseudo-inverses $T_\Psi^\dagger$ and $T_\Phi^\dagger$, initially defined on $\mathcal{M}^\Psi$ and $\mathcal{M}_\Psi$, respectively, can in fact be extended to the entire Hilbert spaces, i.e. $T_\Psi^\dagger : \mathcal{H} \to l^2(I)$ and $T_\Phi^\dagger : \mathcal{K} \to l^2(K)$ and this extension is given explicitly via the dual frame analysis operator.

C.1. **Operator Aliasing and Representation Equivalence.** Once the discretization is chosen – which is equivalent to choosing the frame sequences for input and output spaces (Ψ, Φ) – connecting the continuous operator U with its discrete counterpart u is the notion of operator aliasing. Given any mapping $u : l^2(I) \to l^2(K)$, we can build the operator $T_\Phi \circ u \circ T_\Psi^\dagger : \mathcal{H} \to \mathcal{K}$, whose definition clearly 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 definition of aliasing for functions in Hilbert space, we can define the aliasing error of U relative to the discrete representation u as,

Definition 4 (Operator aliasing). The aliasing error operator $\varepsilon(U, u, \Psi, \Phi) : \mathrm{Dom} \, U \subseteq \mathcal{H} \to \mathcal{K}$ is given by

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

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

An aliasing error of zero implies that the operator U can be perfectly represented by first discretizing the function with $T_\Psi^\dagger$, applying u, and then reconstructing with T_Φ, or equivalently,

that the diagram in Figure 3a commutes, i.e. black and 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 same as accessing the underlying continuous operator U. In practice, the discrete representation u of

(a) An alias-free operator's diagram.

(b) ReNO: alias-free u can be constructed by discr the operator U for different discretization given by Ψ, Φ.

(c) ReNO: discrete representations u, u' for different discretizations are equivalent
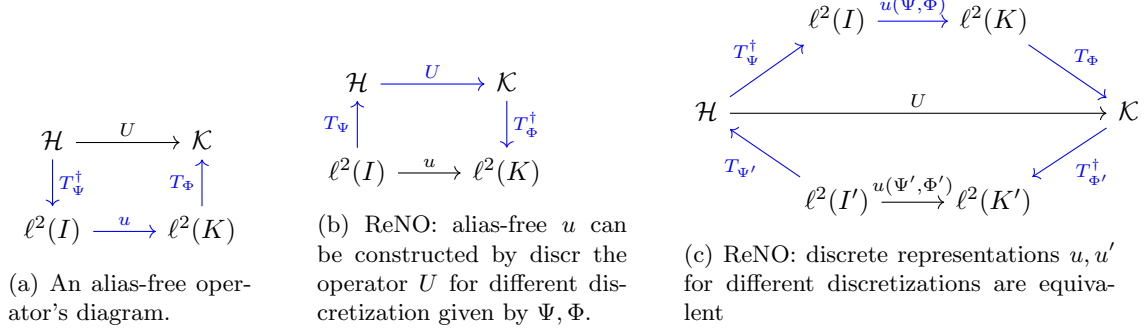
FIGURE 3. Alias-free framework

the operator U depends on the choice of the frame sequences. This means that – as mentioned at the beginning of the section – there is one map for every input/output frame sequence Φ, Ψ. The consistency between operations u, u' with respect to different frame sequences can be evaluated using the following error.

Definition 5 (Representation equivalence error)**.** Suppose that u, u' are discrete maps with associated frame sequences Ψ, Φ and Ψ', Φ', respectively. Then, the representation equivalence error is given by the function $\tau(u, u') : \ell^2(I) \to \ell^2(K)$, defined as:

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

and the corresponding scalar error is $\|\tau(u, u')\|$.

Intuitively, this amount to computing each mapping on their given discretization, and comparing them by expressing u' in the frames associated to u

Remark 6. When there is no aliasing then the representation equivalence error is 0 and the change of frame formula holds (equate $\tau(u, u')$ to 0 to get the change of frame formula.)

C.2. **Representation equivalent Neural Operator (ReNO).** We now introduce the concept of R*Representation equivalent Neural Operator (ReNO)*. To this end, for any pair (Ψ, Φ) of frame sequences for $\mathcal{H}$ and $\mathcal{K}$, we consider a mapping at the discrete level $u(\Psi, \Phi) : \mathrm{Ran} T_\Psi^\dagger \to \mathrm{Ran} T_\Phi^\dagger$, which handles the discrete representations of the functions. This map is discretization dependent and changes with the choices of frame sequences. When this is clear from the context, we will refer to $u(\Psi, \Phi)$ simply as u. See B.1 for the explanation of why $\mathrm{Ran}\, T_\Psi^\dagger$ or $\mathrm{Ran}\, T_\Phi^\dagger$ is choosen as domain and codomain respectively.

Definition 6 (Representation equivalent Neural Operators (ReNO))**.** We say that (U, u) is a ReNO if for every pair (Ψ, Φ) of frame sequences that satisfy $\mathrm{Dom}\, U \subseteq \mathcal{M}_\Psi$ and $\mathrm{Ran}\, U \subseteq \mathcal{M}_\Phi$, there is no aliasing, i.e. the aliasing error operator is identical to zero:

$$\varepsilon(U, u, \Psi, \Phi) = 0 \tag{C.1}$$

We will write this property in short as $\varepsilon(U, u) = 0$.

In other words, the diagram in 3a commutes for every every pair of (Ψ, Φ). In this case, the discrete representations $u(\Psi, \Phi)$ 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. The domain and range conditions in above definition simply imply that the frame can adequately represent input and output functions of U.

Remark 7. If aliasing error $\varepsilon(U, u, \Psi, \Phi) = 0$ then the assumption that $u(\Psi, \Phi)$ maps $\text{Ran } T_\Psi^\dagger \subseteq \ell^2(I)$ into $\text{Ran } T_\Phi^\dagger \subseteq \ell^2(K)$ implies that

$$u(\Psi, \Phi) = T_\Phi^\dagger \circ U \circ T_\Psi. \tag{C.2}$$

We observe that this definition of $u(\Psi, \Phi)$ is such that the diagram in Figure 3b commutes. In other words, once we fix the discrete representation associated to the input and output functions, there exists a unique way to defined a discretization $u(\Psi, \Phi)$ that is consistent with the continuous operator U and this is given by the above equation.

Proof. If aliasing error $\varepsilon(U, u, \Psi, \Phi) = 0$, then

$$U = T_\Phi \circ u(\Psi, \Phi) \circ T_\Psi^\dagger. \tag{C.3}$$

By above equation, we obtain

$$T_\Phi^\dagger \circ U \circ T_\Psi = T_\Phi^\dagger \circ T_\Phi \circ u(\Psi, \Phi) \circ T_\Psi^\dagger \circ T_\Psi = u(\Psi, \Phi).$$

where the last equality follows from the fact that $T_\Psi^\dagger \circ T_\Psi$ is the orthogonal projection onto $(\text{Ker}(T_\Psi))^\perp = \text{Ran}(T_\Psi^\dagger)$ and, by assumption, $u(\Psi, \Phi)$ maps $\text{Ran } T_\Psi^\dagger$ into $\text{Ran } T_\Phi^\dagger$. ∎

In partiuclar, Remark 7 directly implies the formula to go from one discrete representation to another, as

$$u(\Psi', \Phi') = T_{\Psi'}^\dagger \circ T_\Phi \circ u(\Psi, \Phi) \circ T_\Psi^\dagger \circ T_{\Psi'} \tag{C.4}$$

whenever the pairs of frame sequences (Ψ, Φ) and (Ψ', Ψ') satisfy the conditions in Definition 6. In other words, the diagram in 3c commutes.

Proposition 2 (Equivalence of ReNO discrete representations). *Let (U, u) be a ReNO. For any two frame sequence pairs (Ψ, Φ) and (Ψ', Φ') satisfying conditions in Definition 6, we have that*

$$\tau(u, u') = 0,$$

where, by a slight abuse of notation, u' denotes $u(\Psi', \Phi')$.

Proof. Immediately implied by the formula C.4. This proposition establishes the link between aliasing and representation equivalence i.e. if aliasing is 0 then it will imply the representation equivalence error to be 0. ∎

REFERENCES

[1] Francesca Bartolucci, Emmanuel de Bézenac, Bogdan Raonić, Roberto Molinaro, Siddhartha Mishra, and Rima Alaifari. Representation equivalent neural operators: a framework for alias-free operator learning, 2023. URL https://arxiv.org/abs/2305.19913.

[2] Ole Christensen. *Frames and Bases.* Applied and Numerical Harmonic Analysis. Birkhauser Boston, Secaucus, NJ, 2008 edition, July 2008.

[3] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2021. URL https://arxiv.org/abs/2010.08895.

[4] Bogdan Raonić, 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, 2023. URL https://arxiv.org/abs/2302.01178.

[5] M. Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569–587, 2000. doi: 10.1109/5.843002.

[6] Yinhao Zhu and Nicholas Zabaras. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, 366:415–447, August 2018. ISSN 0021-9991. doi: 10.1016/j.jcp.2018.04.018. URL http://dx.doi.org/10.1016/j.jcp.2018.04.018.