<u>Theorem</u>: Universal Approx theorem for operators.

Suppose σ is a nonlinear, continuos, non-polynomial, differentiable (in $C^1(\mathbb{R})$)

X is a Banach Space, $\underbrace{K_1 \subseteq X,\ K_2 \subseteq \mathbb{R}^d}_{\text{Compact}}$, V is a compact set in $C(K_1)$.

If G is a (possibly nonlinear) continuos operator $V \to C(K_2)$, then $\forall \varepsilon > 0$,
$\exists n, p, m \in \mathbb{Z}^+$ and const. $c_i^k, \xi_{ij}^k, \theta_i^k, \xi \in \mathbb{R}$
$x_i \in K_1$, $\omega_k \in \mathbb{R}^d$ for $i = 1, \dots, n$ and $k = 1, \dots, p$, $j = 1, \dots, m$
Such that

$$\left| G(u)(y) - \sum_{k=1}^{p} \sum_{i=1}^{n} c_i^k \, r\left(\sum_{j=1}^{m} \xi_{ij}^k u(y_j) + \theta_i^k \right) r(\omega_k \cdot y + \xi_k) \right| < \varepsilon$$

$\forall u \in V$, $\forall y \in V_2$

Paper 1 (Li et al, 2020) — GKN

<u>Recall</u>: Neural Operators generalize the task of ML by approx Maps btwn inf. dimen Spaces & btwn finite approx of those spaces.

A, U are seperable Banach Spaces

$F^\dagger: A \to U$, possibly nonlinear

$a_j \sim \mu$ s.t $u_j = F^\dagger(a_j) + \varepsilon$, $\varepsilon \sim \overset{(\text{noise})}{\mu'}$

Training Pairs $\{a_j, u_j\}_{j=1}^{N}$ with K observations $a_j|_{P_K}$ & $u_j|_{P_K}$

Want to find $\theta^\dagger \in \Theta$ s.t $F_\theta: A \to U$ $\quad F_{\theta^\dagger} \approx F^\dagger$

$$\theta^\dagger = \underset{\theta \in \Theta}{\operatorname{argmin}} \ \mathbb{E}_{a \sim \mu} \left[\underset{\text{cost fct}}{C}\left(F_\theta(a), F^\dagger(a)\right) \right]$$

Consider PDE of the form $\begin{cases} (L_a u)(x) = f(x) & x \in D \\ u(x) = 0 & \partial D \end{cases}$ $\quad \{a_j, u_j\}$

with Solution $u: D \to \mathbb{R}$ & param $\underset{\text{features}}{a}: D \to \mathbb{R}$

<u>Green's Function</u>: $\underset{}{\overset{\text{Green's Fct}}{G_a}}$ is the unique distributional Solution to

$$(L_a G_a(x, \cdot))(y) = \delta_x = 1_D(x) = \begin{cases} 1 & x \in D \\ 0, & x \notin D \end{cases}$$
$\forall x \in D$

L_a for a PDE w/homogeneous boundary data.

Notice $L_a \int_D G_a(x,y) f(y)\, dy = \int_D L_a G_a(x,y) f(y)\, dy$

Thus $\boxed{u(x) = \int_D G_a(x,y) f(y)\, dy}$ $\quad = \int_D f(y) \delta_x(y)\, dy$
$\quad = \int_D f(y) \delta_x(dy) = f(x)$

Lebesgue Measure Supported on $\overset{\text{Ball}}{B(x,r)}$

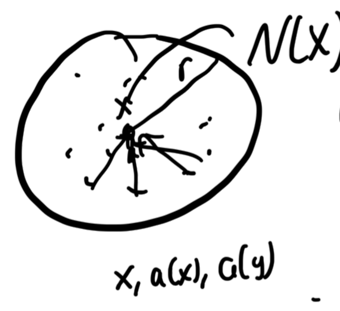
Let the update to representation $v_t \to v_{t+1}$ be

$$v_{t+1} = \sigma\left(W v_t(x) + \int_D K_\phi(x, y, a(x), a(y)) v_t(y)\, \nu_x(dy)\right)$$

$$v_0(x) = P(x, a(x), a_\varepsilon(x), \nabla a_\varepsilon(x)) + p$$

$$v_{t+1}(x) = \sigma\left(W v_t(x) + \int_{B(x,r)} K_\phi(x, y, a(x), a(y)) v_t(y)\, dy\right)$$

$u(x) = Q \underset{\text{final layer}}{v_T}(x) + q$ $\qquad$ Monte Carlo Approx



$N(x) \subseteq B(x,r)$

$$v_{t+1} = \sigma\left(W v_t(x) + \frac{1}{|N(x)|} \sum_{y \in N(x)}^{D} K_\phi(e(x,y)) v_t(y)\right)$$

<u>Fourier Neural Operator</u>

<u>Fast Fourier Transform</u>: If f is Supported on D, and D has been discretized with n points. Now denote $\hat{F}$ as the FFT operation:

$$(\hat{F} f)(k) = \sum_{x_1=0}^{s_1-1} \sum_{x_2=0}^{s_2-1} \cdots \sum_{x_d=0}^{s_d-1} f(x_1, \dots, x_d) e^{-2\pi i \sum_{j=1}^{d} \frac{x_j k_j}{s_j}}$$

$$s_1 \cdot s_2 \cdot \dots \cdot s_d = n$$

$$(\hat{F}^{-1} f) = \sum_{k_1=0}^{s_1-1} \cdots \sum_{k_d=0}^{s_d-1} f(k_1, k_2, \dots, k_d) e^{2\pi i \sum_{j=1}^{d} \frac{x_j k_j}{s_j}}$$

$$v_{t+1} = \sigma\Big(W v_t(x) + \overbrace{\int_D K_\phi(x, y, a(x), a(y)) v_t(y)\, \nu_x(dy)}^{(K(a;\theta) v_t)(x)}\Big) \quad (*)$$

$$v_{t+1}(x) = \sigma\left(W v_t(x) + (K(a;\theta) v_t(x))\right)$$

<u>FNO</u>: $(K(a;\theta) v_t)(x) = \hat{F}^{-1}\left(\hat{F}(K_\phi) \cdot \hat{F}(v_t)\right)(x)$

<u>Convolution Theorem</u>: Product in Frequency $\equiv$ Convolution in time

Why not Stay in frequency domain: $\int v_0 K_\phi^1 K_\phi^2 \dots \quad \hat{F}^{-1}\left((\hat{F} K_\phi^2)(\hat{F} K_\phi^1 (F v_0)(x_0)\right.$

What we not $\cdots K_\phi^3 * (K_\phi^2 * (K_\phi^1 * v_0))$

Graph: $\Theta(n^2)$ $\qquad$ <u>FNO</u>: $\Theta(n \log n)$