Operator learning for multi-patch domains

Massimiliano Ghiotto
PhD student at University of Pavia

Joint work with:

Carlo Marcati, Giancarlo Sangalli

UNIVERSITÀ DI PAVIA

Table of Contents

UNIVERSITÀ DI PAVIA

Definition of Neural Operator

$$\mathcal{N}_\theta : \mathcal{A}(D, \mathbb{R}^{d_a}) \to \mathcal{U}(D, \mathbb{R}^{d_u}), \quad \mathcal{N}_\theta := \mathcal{Q} \circ \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}.$$

1. **Lifting:** linear and local operator

$$\mathcal{R} : \mathcal{A}(D, \mathbb{R}^{d_a}) \to \mathcal{U}(D, \mathbb{R}^{d_{v_1}}), \quad \mathcal{R}(a)(x) = R \cdot a(x), \; R \in \mathbb{R}^{d_{v_1} \times d_a}$$

Definition of Neural Operator

$$\mathcal{N}_\theta : \mathcal{A}(D, \mathbb{R}^{d_a}) \to \mathcal{U}(D, \mathbb{R}^{d_u}), \quad \mathcal{N}_\theta := \mathcal{Q} \circ \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}.$$

1. **Lifting:** linear and local operator

$$\mathcal{R} : \mathcal{A}(D, \mathbb{R}^{d_a}) \to \mathcal{U}(D, \mathbb{R}^{d_{v_1}}), \quad \mathcal{R}(a)(x) = R \cdot a(x), \; R \in \mathbb{R}^{d_{v_1} \times d_a}$$

2. **Integral operator:** for $t = 1, \dots, L$

$$\mathcal{L}_t : \mathcal{U}(D, \mathbb{R}^{d_{v_t}}) \to \mathcal{U}(D, \mathbb{R}^{d_{v_t}})$$

$$\mathcal{L}_t(v)(x) := \sigma\Big(W_t v(x) + b_t(x) + (\mathcal{K}_t(a, \theta)v)(x)\Big)$$

with $\mathcal{K}_t(a, \theta)$ linear and non-local operator.

Definition of Neural Operator

$$\mathcal{N}_\theta : \mathcal{A}(D, \mathbb{R}^{d_a}) \to \mathcal{U}(D, \mathbb{R}^{d_u}), \quad \mathcal{N}_\theta := \mathcal{Q} \circ \mathcal{L}_L \circ \cdots \circ \mathcal{L}_1 \circ \mathcal{R}.$$

1. **Lifting:** linear and local operator

$$\mathcal{R} : \mathcal{A}(D, \mathbb{R}^{d_a}) \to \mathcal{U}(D, \mathbb{R}^{d_{v_1}}), \quad \mathcal{R}(a)(x) = R \cdot a(x), \; R \in \mathbb{R}^{d_{v_1} \times d_a}$$

2. **Integral operator:** for $t = 1, \dots, L$

$$\mathcal{L}_t : \mathcal{U}(D, \mathbb{R}^{d_{v_t}}) \to \mathcal{U}(D, \mathbb{R}^{d_{v_t}})$$

$$\mathcal{L}_t(v)(x) := \sigma\Big(W_t v(x) + b_t(x) + (\mathcal{K}_t(a,\theta)v)(x)\Big)$$

with $\mathcal{K}_t(a, \theta)$ linear and non-local operator.

3. **Projection:** linear and local operator

$$\mathcal{Q} : \mathcal{U}(D_L, \mathbb{R}^{d_{v_L}}) \to \mathcal{U}(D, \mathbb{R}^{d_u}), \quad \mathcal{Q}(v_L)(x) = Q \cdot v_L(x), \; Q \in \mathbb{R}^{d_{v_u} \times d_{v_L}}.$$

Integral operator

There are different ways to define the integral operator $\mathcal{K}_t$:

- defining $\kappa_{t,\theta} \in C(D \times D, \mathbb{R}^{d_{v_t} \times d_{v_t}})$

$$(\mathcal{K}_t(a,\theta)v_t)(x) = (\mathcal{K}_t(\theta)v_t)(x) = \int_D \kappa_{t,\theta}(x,y)v_t(y)\; d\mu_t(y).$$

Integral operator

There are different ways to define the integral operator $\mathcal{K}_t$:

- defining $\kappa_{t,\theta} \in C(D \times D, \mathbb{R}^{d_{v_t} \times d_{v_t}})$

$$(\mathcal{K}_t(a,\theta)v_t)(x) = (\mathcal{K}_t(\theta)v_t)(x) = \int_D \kappa_{t,\theta}(x,y)v_t(y)\ d\mu_t(y).$$

- defining $\kappa_{t,\theta} \in C(D \times D \times \mathbb{R}^{d_a} \times \mathbb{R}^{d_a}, \mathbb{R}^{d_{v_t} \times d_{v_t}})$

$$(\mathcal{K}_t(a,\theta)v_t)(x) = \int_D \kappa_{t,\theta}\left(x, y, a\left(x\right), a\left(y\right)\right) v_t(y)\ d\mu_t(y).$$

Integral operator

1 Operator Learning

There are different ways to define the integral operator $\mathcal{K}_t$:

- defining $\kappa_{t,\theta} \in C(D \times D, \mathbb{R}^{d_{v_t} \times d_{v_t}})$

$$(\mathcal{K}_t(a,\theta)v_t)(x) = (\mathcal{K}_t(\theta)v_t)(x) = \int_D \kappa_{t,\theta}(x,y)v_t(y)\ d\mu_t(y).$$

- defining $\kappa_{t,\theta} \in C(D \times D \times \mathbb{R}^{d_a} \times \mathbb{R}^{d_a}, \mathbb{R}^{d_{v_t} \times d_{v_t}})$

$$(\mathcal{K}_t(a,\theta)v_t)(x) = \int_D \kappa_{t,\theta}\left(x, y, a\left(x\right), a\left(y\right)\right) v_t(y)\ d\mu_t(y).$$

- defining $\kappa_{t,\theta} \in C(D \times D \times \mathbb{R}^{d_{v_t}} \times \mathbb{R}^{d_{v_t}}, \mathbb{R}^{d_{v_t} \times d_{v_t}})$

$$(\mathcal{K}_t(a,\theta)v_t)(x) = \int_D \kappa_{t,\theta}\left(x, y, v_t\left(x\right), v_t(y)\right) v_t(y)\ d\mu_t(y).$$

Table of Contents

UNIVERSITÀ DI PAVIA

Definition of Fourier Neural Operator (FNO)

For defining the Fourier Neural Operator we make the first assumption and the further assumption that $\kappa_{t,\theta}(x,y) = \kappa_{t,\theta}(x-y)$,

$$(\mathcal{K}_t(a,\theta)v)(x) = \int_{\mathbb{T}^d} \kappa_{t,\theta}(x-y)v(y)\ dy = (\kappa_{t,\theta} * v)(x).$$

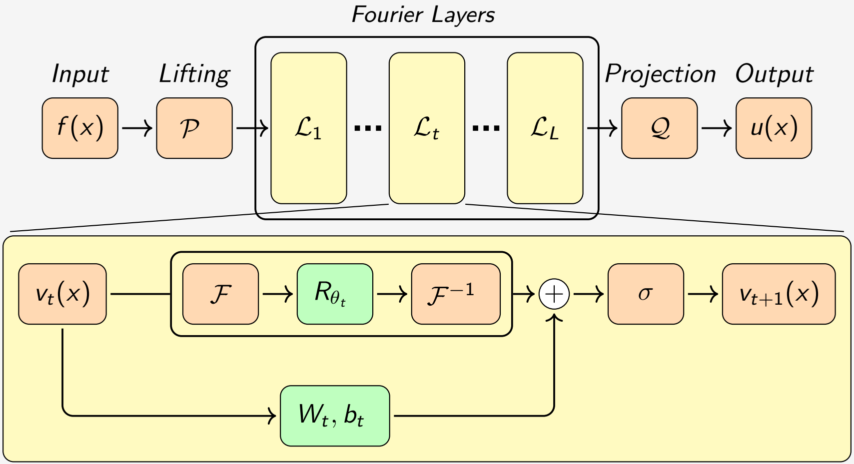
Using the convolution theorem we have

$$(\kappa_{t,\theta} * v)(x) = \mathcal{F}^{-1}\left(\mathcal{F}(\kappa_{t,\theta})(k) \cdot \mathcal{F}(v)(k)\right)(x),$$

and parameterizing $\mathcal{F}(\kappa_{t,\theta})$ with the parameters $P_\theta(k) \in \mathbb{C}^{d_v \times d_v}$ $\forall k$ we have

$$(\mathcal{K}_t(a,\theta)v)(x) = \mathcal{F}^{-1}\left(P_\theta(k) \cdot \mathcal{F}(v)(k)\right)(x)$$

Fourier Neural Operator



Teorema di approssimazione universale per le FNO

Siano $s, s' \geq 0$ e

$$\mathcal{G} : H^s(\mathbb{T}^d, \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d, \mathbb{R}^{d_u})$$

un operatore continuo. Siano $K \subset H^s(\mathbb{T}^d, \mathbb{R}^{d_a})$ un insieme compatto e $\sigma \in \mathbb{C}^\infty(\mathbb{R})$ funzione di attivazione non polinomiale e globalmente Lipschitz. Allora, per ogni $\varepsilon > 0$, esiste un operatore continuo con struttura data da una FNO

$$\mathcal{N} : H^s(\mathbb{T}^d, \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d, \mathbb{R}^{d_u})$$

tale che:

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}(a)\|_{H^{s'}} \leq \varepsilon.$$

Pseudo Operatori Neurali di Fourier (ψ-FNO)

Uno pseudo-operatore di Fourier è una mappa

$$\mathcal{N}^* : \mathcal{A}(\mathbb{T}^d, \mathbb{R}^{d_a}) \to \mathcal{U}(\mathbb{T}^d, \mathbb{R}^{d_u}), \qquad a \mapsto \mathcal{N}^*(a),$$

della forma

$$\mathcal{N}^*(a) = \mathcal{Q} \circ I_N \circ \mathcal{L}_L \circ I_N \circ \cdots \circ \mathcal{L}_1 \circ I_N \circ \mathcal{R}(a),$$

dove I_N denota la proiezione pseudo-spettrale di Fourier di grado N

$$I_N : C(\mathbb{T}^d) \to L^2_N(\mathbb{T}^d), \quad u \mapsto I_N u.$$

Uno ψ-FNO si può identificare con una mappa finita dimensionale

$$\widetilde{\mathcal{N}}^* : \mathbb{R}^{d_a \times \mathcal{I}_N} \to \mathbb{R}^{d_a \times \mathcal{I}_N}, \quad \widetilde{\mathcal{N}}^* : a \mapsto \widetilde{\mathcal{N}}^*(a)$$

$$\widetilde{\mathcal{N}}^*(a)_j = \mathcal{N}^*(a)(x_j)$$

Teorema di approssimazione universale per le ψ-FNO

Siano $s > d/2$, $s' \geq 0$ e

$$\mathcal{G} : H^s(\mathbb{T}^d, \mathbb{R}^{d_a}) \to H^{s'}(\mathbb{T}^d, \mathbb{R}^{d_u})$$

un operatore continuo. Siano $K \subset H^s(\mathbb{T}^d, \mathbb{R}^{d_a})$ un insieme compatto e $\sigma \in \mathbb{C}^\infty(\mathbb{R})$ una funzione di attivazione non polinomiale e globalmente Lipschitz. Allora, per ogni $\varepsilon > 0$, esiste un $N \in \mathbb{N}$ tale che la ψ-FNO

$$\mathcal{N}^* : L^2_N(\mathbb{T}^d, \mathbb{R}^{d_a}) \to L^2_N(\mathbb{T}^d, \mathbb{R}^{d_u})$$

soddisfa:

$$\sup_{a \in K} \|\mathcal{G}(a) - \mathcal{N}^*(a)\|_{H^{s'}} \leq \varepsilon.$$

Teorema

Sia $s > d/2$, $\lambda \in (0,1)$ e consideriamo l'operatore soluzione del problema di Darcy

$$\mathcal{G} : \mathcal{A}^s_\lambda(\mathbb{T}^d) \to H^1(\mathbb{T}^d).$$

Fissata $\sigma \in C^3(\mathbb{R})$ non polinomiale per ogni $N \in \mathbb{N}$ esiste $C > 0$ e una ψ-FNO

$$\mathcal{N}^* : \mathcal{A}^s_\lambda(\mathbb{T}^d) \to H^1(\mathbb{T}^d)$$

tale che

$$\sup_{a \in \mathcal{A}^s_\lambda} \|\mathcal{G}(a) - \mathcal{N}^*(a)\|_{H^1(\mathbb{T}^d)} \leq C N^{-k}$$

e $\mathrm{depth}(\mathcal{N}^*) \leq C \log(N)$, $\mathrm{lift}(\mathcal{N}^*) \leq C$, $\mathrm{size}(\mathcal{N}^*) \lesssim N^d \log(N)$.

Table of Contents

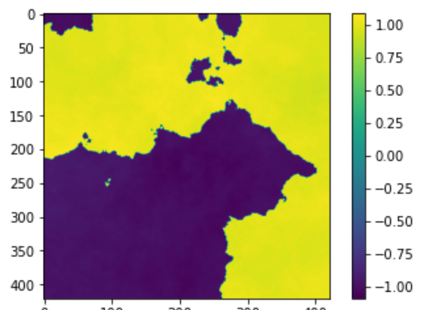
UNIVERSITÀ DI PAVIA

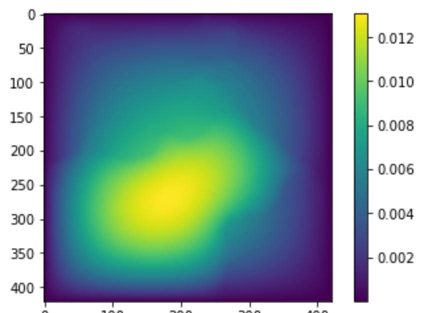
Problema di Darcy

$$\begin{cases} -\nabla(a \cdot \nabla u) = f, & \text{in } D \\ u = 0, & \text{on } \partial D \end{cases}$$

con $D = [0,1]^2$, $\mathcal{A} = L^\infty(D, \mathbb{R}^+)$, $\mathcal{U} = H_0^1(D, \mathbb{R})$ e $f \equiv 1$.

$$\mathcal{G} : L^\infty(D, \mathbb{R}^+) \to H_0^1(D, \mathbb{R}), \qquad \mathcal{G} : a \mapsto u.$$



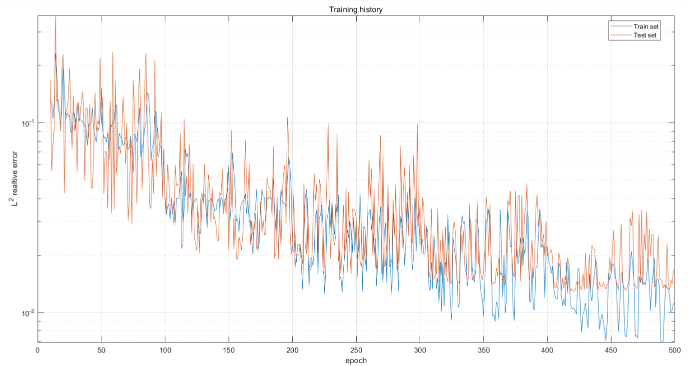


Norma relativa L^2

$$\left\| \frac{\mathcal{G} - \mathcal{N}_\theta^*}{\mathcal{G}} \right\|_{L^2_\mu(L^\infty, L^2)} = \mathbb{E}_{a\sim\mu} \frac{\|\mathcal{G}(a) - \mathcal{N}_\theta^*(a)\|^2_{L^2(D)}}{\|\mathcal{G}(a)\|^2_{L^2(D)}} \approx$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{M} \left| u^{(i)}(x_k) - \widetilde{\mathcal{N}}_\theta^*\left(a^{(i)}\right)(x_k) \right|^2}{\sum_{k=1}^{M} \left| u^{(i)}(x_k) \right|^2} \right)$$

con $D_k = \{x_k\}_{k=1}^M \subset D = [0,1]$ e $M = 85^2$. Dataset $\left\{a^{(i)}, u^{(i)}\right\}_{i=1}^N$ con $a^{(i)} \sim \mu = T_\# N(0, C)$ i.i.d. e $u^{(i)} = \mathcal{G}(a^{(i)})$ soluzione approssimata e con valutazioni puntuali $\left\{a^{(i)}_{|D_k}, u^{(i)}_{|D_k}\right\}_{i=1}^N$.



Griglia 85×85 con $L = 4$, $d_v = 32$, $k_{max} = 12$, $\sigma = ReLU$, 1000 funzioni per l'allenamento e 200 per il test, 500 epoche e learning rate inizializzato a $0,001$ e dimezzato ogni 100 epoche.

train error	rel. error L^2	parameters	training time
0.01305	0.01804	2 363 681	6 hours

Norma relativa H^1

$$\left\| \frac{\mathcal{G} - \mathcal{N}_\theta^*}{\mathcal{G}} \right\|_{L^2_\mu(L^\infty, H^1_0)} = \mathbb{E}_{a\sim\mu} \frac{\|\mathcal{G}(a) - \mathcal{N}_\theta^*(a)\|^2_{H^1(D)}}{\|\mathcal{G}(a)\|^2_{H^1(D)}},$$

dove

$$\|f\|^2_{H^1(\mathbb{T}^d)} = \sum_{k\in\mathbb{Z}^d} \left(1 + |k|^2\right)^s \left|\widehat{f}(k)\right|^2$$

da cui

$$\left\| \frac{\mathcal{G} - \mathcal{N}_\theta^*}{\mathcal{G}} \right\|_{L^2_\mu(L^\infty, H^1_0)} \approx$$

$$\approx \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{k\in\mathbb{Z}_N} \left(1 + |k|^2\right)^s \left|\widehat{u^{(i)}}(k) - \widehat{\mathcal{N}_\theta^*\left(a^{(i)}\right)}(k)\right|^2}{\sum_{k\in\mathbb{Z}_N} \left(1 + |k|^2\right)^s \left|\widehat{u^{(i)}}(k)\right|^2}$$

La norma relativa H^1 come funziona di perdita aiuta l'allenamento dell'operatore neurale.

funzione di perdita	errore rel. L^2	errore rel. H^1
rel. L^2	0.01804	0.06944
rel. H^1	0.01203	0.04793

Nuove prestazioni con la norma relativa H^1 come funzione di perdita.

funzione di perdita	errore rel. L^2	errore rel. H^1
rel. L^2	0.01038	0.05979
rel. H^1	**0.007220**	**0.03803**

Entrambe le architetture hanno 2 376 449 parametri ed impiegano 7 ore per l'allenamento.

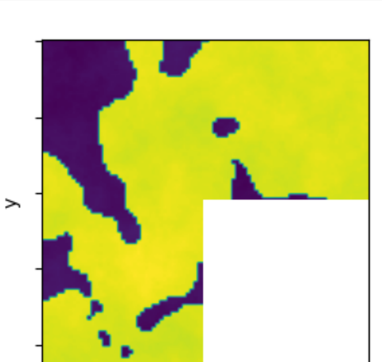
Table of Contents

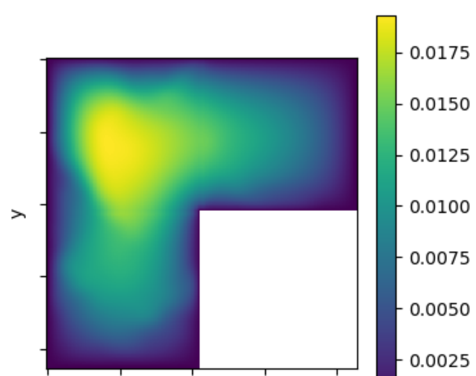
UNIVERSITÀ DI PAVIA

Problema di Darcy su un dominio ad L

$$\begin{cases} -\nabla(a \cdot \nabla u) = f, & \text{in } \Omega \\ u = 0, & \text{in } \partial\Omega \end{cases}$$

problema di Darcy su $\Omega = [-1,1]^2 \setminus (0,1) \times (-1,0)$, $f \equiv 1$.



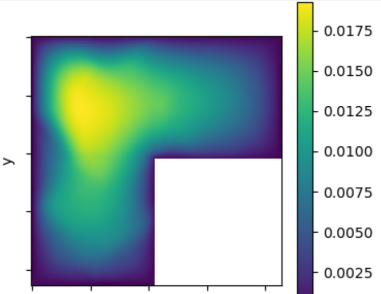


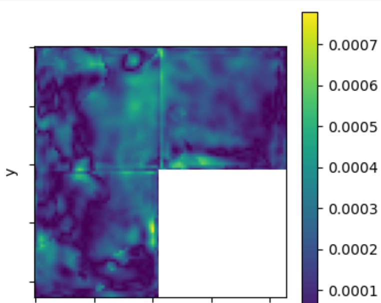
Fourier continuation

Massimiliano Ghiotto
UNIVERSITÀ DI PAVIA

Gli operatori neurali di Fourier hanno la limitazione che sono ristretti a domini rettangolari. Quando ho un dominio irregolare posso estenderlo a un dominio rettangolare più grande, la funzione di perdita calcolata solo sul dominio originale.

errore rel. L^2	parametri	tempo allenamento
0.02450	2 363 681	7 hours





Operator learningfor multi-patch domains

Thank you for listening!

UNIVERSITÀ DI PAVIA

Trasformata di Fourier

- Sia $v \in L^2(\mathbb{T}^d)$, la trasformata di Fourier è

$$\begin{aligned} \mathcal{F} : L^2(\mathbb{T}^d, \mathbb{C}^n) &\to \ell^2(\mathbb{Z}^d, \mathbb{C}^n) \\ v &\mapsto \mathcal{F}(v) \end{aligned}$$

$$\mathcal{F}(v)(k) := \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} v(x) e^{-i\langle k, x\rangle} \, dx, \quad \forall k \in \mathbb{Z}^d.$$

- Data $\widehat{v} = \{\widehat{v}_k\}_{k \in \mathbb{Z}^d} \in \ell^2(\mathbb{Z}^d, \mathbb{C}^n)$, la trasformata inversa di Fourier è

$$\begin{aligned} \mathcal{F}^{-1} : \ell^2(\mathbb{Z}^d, \mathbb{C}^n) &\to L^2(D, \mathbb{C}^n) \\ \widehat{v} &\mapsto \mathcal{F}^{-1}(\widehat{v}) \end{aligned}$$

$$(\mathcal{F}^{-1}\widehat{v})(x) = \sum_{k \in \mathbb{Z}^d} \widehat{v}_k e^{i\langle k, x\rangle} \qquad \forall x \in D.$$

Trasformata discreta di Fourier

Sia $N \in \mathbb{N}$ e fissata una griglia regolare $\{x_j\}_{j\in\mathcal{I}_N}$ con $x_j = (2\pi j)/(2N+1) \in \mathbb{T}^d$, $j \in \mathcal{I}_N = \{0, \dots, 2N\}^d$ e scelto un insieme per i modi di Fourier $\mathcal{K}_N := \{k \in \mathbb{Z}^d : |k|_\infty \leq N\}$. Definiamo la trasformata discreta di Fourier come

$$\mathcal{F}_N : \mathbb{R}^{\mathcal{I}_N} \to \mathbb{C}^{\mathcal{K}_N}$$

$$\mathcal{F}_N(v)(k) := \frac{1}{(2N+1)^d} \sum_{j\in\mathcal{I}_N} v(x_j) e^{-2\pi i\langle j,k\rangle/N}, \quad \forall k \in \mathcal{K}_N,$$

e la trasformata inversa discreta di Fourier come

$$\mathcal{F}_N^{-1} : \mathbb{C}^{\mathcal{K}_N} \to \mathbb{R}^{\mathcal{I}_N}$$

$$\mathcal{F}_N^{-1}(\widehat{v})(j) := \sum_{k\in\mathcal{K}_N} \widehat{v}_k e^{2\pi i\langle j,k\rangle/N}, \qquad \forall j \in \mathcal{J}_N.$$

Schema dimostrazione teo. universale FNO

- Proiezione sullo spazio dei polinomi trigonometrici

$$P_N : L^2(\mathbb{T}^d) \to L^2_N(\mathbb{T}^d),$$

$$P_N \left(\sum_{k\in\mathbb{Z}^d} c_k e^{i\langle x,k\rangle} \right) = \sum_{|k|_\infty \le N} c_k e^{i\langle x,k\rangle}, \qquad \forall (c_k)_{k\in\mathbb{Z}^d} \in \ell^2(\mathbb{Z}^d).$$

- Se il teorema universale vale per $s' = 0$ allora vale per per qualsiasi valore di $s' \ge 0$.
- Fissiamo $s' = 0$

$$\mathcal{G}_N : H^s(\mathbb{T}^d, \mathbb{R}^{d_a}) \to L^2(\mathbb{T}^d, \mathbb{R}^{d_u}), \qquad \mathcal{G}_N(a) := P_N \mathcal{G}(P_N a),$$

vale che $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ si ha

$$\|\mathcal{G}(a) - \mathcal{G}_N(a)\|_{L^2} \le \varepsilon, \qquad \forall a \in K.$$

Definiamo l'operatore

$$\widehat{\mathcal{G}}_N : \mathbb{C}^{\mathcal{K}_N} \to \mathbb{C}^{\mathcal{K}_N}, \qquad \widehat{\mathcal{G}}_N(\widehat{a}_k) := \mathcal{F}_N(\mathcal{G}_N(Re(\mathcal{F}_N^{-1}(\widehat{a}_k)))),$$

per il quale vale l'identità

$$\mathcal{G}_N(a) = \mathcal{F}_N^{-1} \circ \widehat{\mathcal{G}}_N \circ \mathcal{F}_N(P_N a),$$

per le funzioni $a \in L^2(\mathbb{T}^d, \mathbb{R}^{d_a})$. Ci si riconduce a dimostrare che gli operatori neurali di Fourier possono approssimare

$$\mathcal{F}_N^{-1}, \ \widehat{\mathcal{G}}_N, \ \mathcal{F}_N(P_N a).$$

Definizione MLP

Let $d \in \mathbb{N}$ and $L \in \mathbb{N}$ with $L \geq 2$ and $\sigma : \mathbb{R} \to \mathbb{R}$ an activation function. Let $A_\ell \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$, $b_\ell \ in \mathbb{R}^\ell$ with $n_\ell \in \mathbb{N}$ for $\ell = 1, \dots, L$ and $n_0 = d$. We call multilayer perceptron (MLP) the function defined as

$$\begin{cases} x_L = A_L x_{L-1} + b_L \\ x_\ell = \sigma\left(A_\ell x_{\ell-1} + b_\ell\right) \end{cases},$$

where x_0 is the input and x_L is the output of the function.

Universal approximation theorem for operator

Suppose that $\sigma \in TW$, X is a Banach space, $K_1 \subset X$, $K_2 \subset \mathbb{R}^d$ are two compact sets in X and $\mathbb{R}^d$ respectively, and V is a compact set in $C(K_1)$. Let G a nonlinear continuous operator which maps V into $C(K_2)$, then for any $\varepsilon > 0$, there are a positive integers n, p, m; real constants c_i^k, θ_i^k, ξ_{ij}^k, $\zeta^k \in \mathbb{R}$, points $w^k \in \mathbb{R}^d$ and $x_j \in K_1$, with $i = 1, \dots, n$, $j = 1, \dots, m$ and $k = 1, \dots, p$, such that

$$\left| G(u)(y) - \sum_{k=1}^{p} \sum_{i=1}^{n} c_i^k \sigma \left(\sum_{j=1}^{m} \xi_{ij}^k u(x_j) + \theta_i^k \right) \sigma(w^k \cdot y + \zeta^k) \right| < \varepsilon$$

holds for all $u \in V$ and $y \in K_2$.