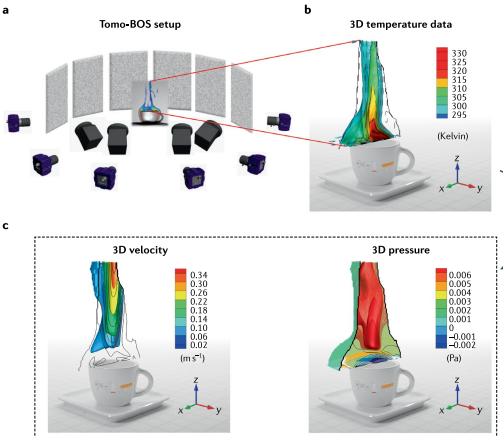
Neural and operator network approximations for elliptic PDEs

Carlo Marcati

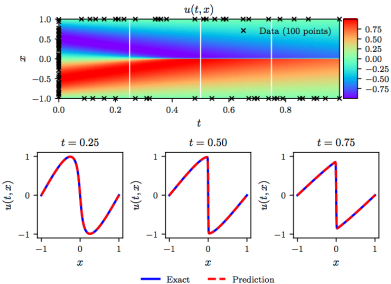
Dipartimento di Matematica
Università di Pavia

UNIVERSITÀ
DI PAVIA

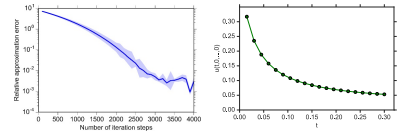
Neural Networks and PDEs



Flow reconstruction [Karniadakis et al., Nature (2021)]



Burgers [Raissi et al., JCP (2019)]



100-dim. Allen-Cahn [Han et al., PNAS (2018)]

Neural networks

Fully connected NN with $L \in \mathbb{N}$ layers and input dim. d

Sequence of matrix-vector couples

$$\Phi = ((A_1, b_1), \dots, (A_L, b_L))$$

with weights $A_\ell \in \mathbb{R}^{n_{\ell-1} \times n_\ell}$ and biases $b_\ell \in \mathbb{R}^{n_\ell}$, for $\ell = 1, \dots, L$ ($n_0 = d$).

Degrees of freedom:

$$\mathrm{depth}(\Phi) := L \qquad \mathrm{size}(\Phi) := \sum_{\ell=1}^{L} (\|A_\ell\|_0 + \|b_\ell\|_0)$$

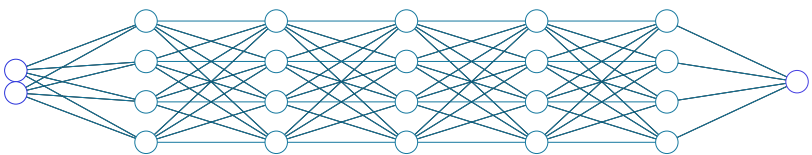
Neural networks

Fully connected NN with $L \in \mathbb{N}$ layers and input dim. d

Sequence of matrix-vector couples

$$\Phi = ((A_1, b_1), \dots, (A_L, b_L))$$

with weights $A_\ell \in \mathbb{R}^{n_{\ell-1} \times n_\ell}$ and biases $b_\ell \in \mathbb{R}^{n_\ell}$, for $\ell = 1, \dots, L$ ($n_0 = d$).



Degrees of freedom:

$$\mathrm{depth}(\Phi) := L \qquad \mathrm{size}(\Phi) := \sum_{\ell=1}^{L} (\|A_\ell\|_0 + \|b_\ell\|_0)$$

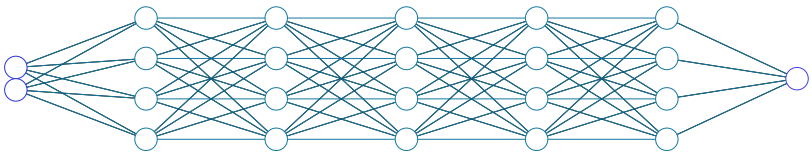
Neural networks

Fully connected NN with $L \in \mathbb{N}$ layers and input dim. d

Sequence of matrix-vector couples

$$\Phi = ((A_1, b_1), \dots, (A_L, b_L))$$

with weights $A_\ell \in \mathbb{R}^{n_{\ell-1} \times n_\ell}$ and biases $b_\ell \in \mathbb{R}^{n_\ell}$, for $\ell = 1, \dots, L$ ($n_0 = d$).



Degrees of freedom:

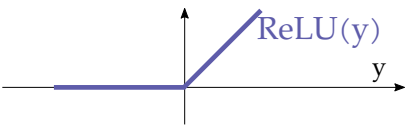
$$\mathbf{depth}(\Phi) := L \qquad \mathbf{size}(\Phi) := \sum_{\ell=1}^{L} (\|A_\ell\|_0 + \|b_\ell\|_0)$$

Neural networks

Realization with ReLU activation

$$\mathbb{R}^d \ni x \mapsto \mathrm{R}(\Phi)(x) = (A_L \cdot + b_L) \circ \rho \circ (A_{L-1} \cdot + b_{L-1}) \circ \cdots \circ \rho \circ (A_1 x + b)$$

with $\rho(y) = \mathrm{ReLU}(y) = \max(y, 0)$ (componentwise)



Some observations

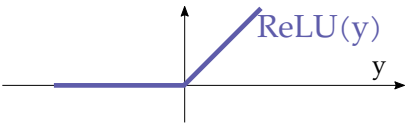
- A ReLU NN is a piecewise linear function
- Every piecewise linear function can be represented as a ReLU NN [Arora et al., ICLR (2018)]
- Automatic differentiation: cheap to compute (exact) derivatives

Neural networks

Realization with ReLU activation

$$\mathbb{R}^d \ni x \mapsto \mathrm{R}(\Phi)(x) = (A_L \cdot + b_L) \circ \rho \circ (A_{L-1} \cdot + b_{L-1}) \circ \cdots \circ \rho \circ (A_1 x + b)$$

with $\rho(y) = \mathrm{ReLU}(y) = \max(y, 0)$ (componentwise)



Some observations

- A ReLU NN is a piecewise linear function
- Every piecewise linear function can be represented as a ReLU NN [Arora et al., ICLR (2018)]
- Automatic differentiation: cheap to compute (exact) derivatives

Operator networks

- Mapping between infinite dimensional spaces $\mathcal{X}$ and $\mathcal{Y}$

$$\mathcal{X} \xrightarrow{\text{Encoding}} \mathbb{R}^n \xrightarrow{\text{Approximation}} \mathbb{R}^m \xrightarrow{\text{Reconstruction}} \mathcal{Y}$$

- Approximation and reconstruction involve NNs

Outline

Theoretical analysis of NNs and ONs for PDEs:

Approximation rates | Convergence of training algorithms

Approximation rates for

1. NN emulation of functions with isolated singularities
2. ON solution of smooth elliptic problems

Outline

Theoretical analysis of NNs and ONs for PDEs:

Approximation rates | Convergence of training algorithms

Approximation rates for

1. NN emulation of functions with isolated singularities
2. ON solution of smooth elliptic problems

NN approximation rates

NN approximation rates

Goal: Quantitative approximation rates

Given a class of functions $\mathcal{A}$ and a norm $\|\cdot\|_X$, for any $\varepsilon \in (0,1)$, find $L(\varepsilon), M(\varepsilon)$ such that

$$\sup_{u\in\mathcal{A}} \inf_{\Phi\in\mathcal{B}_{L(\varepsilon),M(\varepsilon)}} \|u - \mathrm{R}(\Phi)\|_X \leq \varepsilon$$

where

$$\mathcal{B}_{L,M} = \{\Phi \in NN : \mathbf{depth}(\Phi) \leq L,\ \mathbf{size}(\Phi) \leq M\}$$

- We look for uniform rates over $\mathcal{A}$
- Results for PDEs: we want the energy norm $\|\cdot\|_X$

NN approximation rates

Goal: Quantitative approximation rates

Given a class of functions $\mathcal{A}$ and a norm $\|\cdot\|_X$, for any $\varepsilon \in (0,1)$, find $L(\varepsilon), M(\varepsilon)$ such that

$$\sup_{u\in\mathcal{A}} \inf_{\Phi\in\mathcal{B}_{L(\varepsilon),M(\varepsilon)}} \|u - \mathrm{R}(\Phi)\|_X \leq \varepsilon$$

where

$$\mathcal{B}_{L,M} = \{\Phi \in NN : \mathbf{depth}(\Phi) \leq L,\ \mathbf{size}(\Phi) \leq M\}$$

- We look for uniform rates over $\mathcal{A}$
- Results for PDEs: we want the energy norm $\|\cdot\|_X$

Some results in NN approximation theory

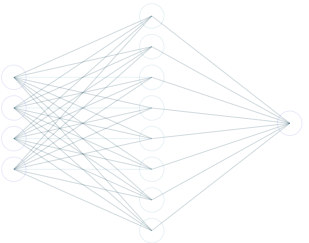
Universal approximation theorem
[Cybenko, Math. Control Signals Systems (1989), Hornik et al., Neural Netw. (1989)]

Let σ be any continuous sigmoidal function. Then, finite sums of the form

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j \cdot x + \theta_j), \qquad y_j, \theta_j \in \mathbb{R}^d,\ \alpha_j \in \mathbb{R}$$

are dense in $C((0,1)^d)$.

- Shallow network: $L = 2$
- $\mathcal{A} \subset C((0,1)^d)$
- $X = L^\infty((0,1)^d)$
- $M(\varepsilon) = ?$



Mathscinet:
Citations
From References: 294
From Reviews: 3

Google Scholar:
Cited by 15966

Some results in NN approximation theory

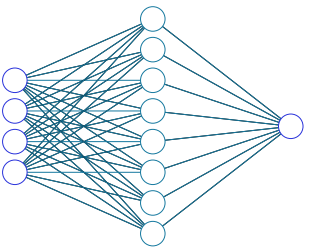
Universal approximation theorem
[Cybenko, Math. Control Signals Systems (1989), Hornik et al., Neural Netw. (1989)]

Let σ be any continuous sigmoidal function. Then, finite sums of the form

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j \cdot x + \theta_j), \qquad y_j, \theta_j \in \mathbb{R}^d,\ \alpha_j \in \mathbb{R}$$

are dense in $C((0,1)^d)$.

- Shallow network: $L = 2$
- $\mathcal{A} \subset C((0,1)^d)$
- $X = L^\infty((0,1)^d)$
- $M(\varepsilon) = ?$



Mathscinet:
Citations
From References: 294
From Reviews: 3

Google Scholar:
Cited by 15966

Some results in NN approximation theory

Universal approximation theorem

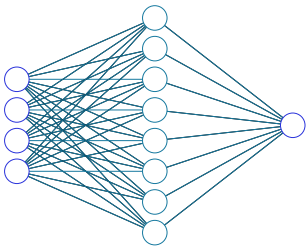
[Cybenko, Math. Control Signals Systems (1989), Hornik et al., Neural Netw. (1989)]

Let σ be any continuous sigmoidal function. Then, finite sums of the form

$$G(x) = \sum_{j=1}^{N} \alpha_j \sigma(y_j \cdot x + \theta_j), \qquad y_j, \theta_j \in \mathbb{R}^d,\ \alpha_j \in \mathbb{R}$$

are dense in $C((0,1)^d)$.

- Shallow network: $L = 2$
- $\mathcal{A} \subset C((0,1)^d)$
- $X = L^\infty((0,1)^d)$
- $M(\varepsilon) = ?$



Mathscinet:

Citations
From References: 294
From Reviews: 3

Google Scholar:

Cited by 15966

Shallow networks

"Universal approximation bounds" [Barron, IEEE Trans. Inf. Theory (1993)]

Sigmoidal σ, $\mathcal{A}$ set of functions with finite integrals of Fourier transforms, any choice of measure $\|\cdot\|_X$ in bounded sets in $\mathbb{R}^d$. Then, $L = 2$ and

$$M(\varepsilon) = \mathcal{O}(\varepsilon^{-2}).$$

- Shallow network
- Dimension independent (except the hidden constant)
- Better than n-widths

Shallow networks

"Universal approximation bounds" [Barron, IEEE Trans. Inf. Theory (1993)]

Sigmoidal σ, $\mathcal{A}$ set of functions with finite integrals of Fourier transforms, any choice of measure $\|\cdot\|_X$ in bounded sets in $\mathbb{R}^d$. Then, $L = 2$ and

$$M(\varepsilon) = \mathcal{O}(\varepsilon^{-2}).$$

- Shallow network
- Dimension independent (except the hidden constant)
- Better than n-widths

Deep ReLU networks

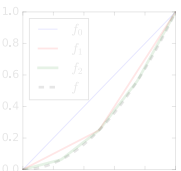
ReLU network approximation in Sobolev spaces [Yarotsky, Neural Netw. (2017)]

ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A} \subset W^{n,\infty}((0,1)^d)$. Then,

$$L(\varepsilon) = \mathcal{O}(|\log(\varepsilon)|), \qquad M(\varepsilon) = \mathcal{O}(\varepsilon^{-d/n}\,|\log(\varepsilon)|).$$

A crucial step:

The function $f(x) = x^2$ on the segment $[0,1]$ can be approximated with any error $\varepsilon > 0$ by a ReLU network having the depth and the number of weights and computation units $\mathcal{O}(|\log \varepsilon|)$.



Deep ReLU networks

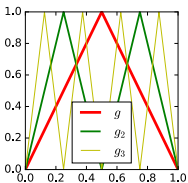
ReLU network approximation in Sobolev spaces [Yarotsky, Neural Netw. (2017)]

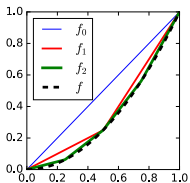
ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A} \subset W^{n,\infty}((0,1)^d)$. Then,

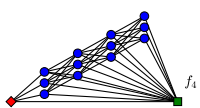
$$L(\varepsilon) = \mathcal{O}(|\log(\varepsilon)|), \qquad M(\varepsilon) = \mathcal{O}(\varepsilon^{-d/n}\,|\log(\varepsilon)|).$$

A crucial step:

The function $f(x) = x^2$ on the segment $[0,1]$ can be approximated with any error $\varepsilon > 0$ by a ReLU network having the depth and the number of weights and computation units $\mathcal{O}(|\log \varepsilon|)$.







"Deeper" ReLU network approximation in Sobolev spaces [Yarotsky & Zhevnerchuk, NeurIPS (2020)]

ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A} \subset W^{n,\infty}((0,1)^d)$. Then,

$$L(\varepsilon) = \mathcal{O}(\varepsilon^{-d/2n}/|\log(\varepsilon)|), \qquad M(\varepsilon) = \mathcal{O}(\varepsilon^{-d/2n}).$$

- Going beyond $-d/n$ requires discontinuous weight assignment,
- Rates higher than $-d/2n$ are not achievable with ReLU networks,
- But for smaller $\mathcal{A}$:

Exponential convergence for analytic functions [E & Wang, Science China Math. (2018)]

ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A}$ subset of analytic functions in $[0,1]^d$. Then,

$$L(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d}), \qquad M(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d})$$

"Deeper" ReLU network approximation in Sobolev spaces [Yarotsky & Zhevnerchuk, NeurIPS (2020)]

ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A} \subset W^{n,\infty}((0,1)^d)$. Then,

$$L(\varepsilon) = \mathcal{O}(\varepsilon^{-d/2n}/|\log(\varepsilon)|), \qquad M(\varepsilon) = \mathcal{O}(\varepsilon^{-d/2n}).$$

- Going beyond $-d/n$ requires discontinuous weight assignment,
- Rates higher than $-d/2n$ are not achievable with ReLU networks,
- But for smaller $\mathcal{A}$:

Exponential convergence for analytic functions [E & Wang, Science China Math. (2018)]

ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A}$ subset of analytic functions in $[0,1]^d$. Then,

$$L(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d}), \qquad M(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d})$$

"Deeper" ReLU network approximation in Sobolev spaces [Yarotsky & Zhevnerchuk, NeurIPS (2020)]

ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A} \subset W^{n,\infty}((0,1)^d)$. Then,

$$L(\varepsilon) = \mathcal{O}(\varepsilon^{-d/2n}/|\log(\varepsilon)|), \qquad M(\varepsilon) = \mathcal{O}(\varepsilon^{-d/2n}).$$

- Going beyond $-d/n$ requires discontinuous weight assignment,
- Rates higher than $-d/2n$ are not achievable with ReLU networks,
- But for smaller $\mathcal{A}$:

Exponential convergence for analytic functions [E & Wang, Science China Math. (2018)]

ReLU activation, $X = L^\infty((0,1)^d)$, $\mathcal{A}$ subset of analytic functions in $[0,1]^d$. Then,

$$L(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d}), \qquad M(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d})$$

Almost everything so far in L^∞ norm, but:

Approximation in Sobolev norms [Gühring et al., Anal. Appl. (2020)]

ReLU activation, $X = W^{s,p}((0,1)^d)$, $\mathcal{A} \subset W^{m,p}((0,1)^d)$. Then,

$$L(\varepsilon) = \mathcal{O}(\varepsilon^{-n/(n-s)}), \qquad M(\varepsilon) = \mathcal{O}(\varepsilon^{-d/(n-s)} \log(\varepsilon^{-n/(n-s)}))$$

When elliptic PDEs have smooth coefficients (or potentials)

- their solutions can have isolated singularities due to the domain
- low regularity in Sobolev spaces, but analytic outside the singular set

Almost everything so far in L^∞ norm, but:

Approximation in Sobolev norms [Gühring et al., Anal. Appl. (2020)]

ReLU activation, $X = W^{s,p}((0,1)^d)$, $\mathcal{A} \subset W^{m,p}((0,1)^d)$. Then,

$$L(\varepsilon) = \mathcal{O}(\varepsilon^{-n/(n-s)}), \qquad M(\varepsilon) = \mathcal{O}(\varepsilon^{-d/(n-s)} \log(\varepsilon^{-n/(n-s)}))$$

When elliptic PDEs have smooth coefficients (or potentials)

- their solutions can have isolated singularities due to the domain
- low regularity in Sobolev spaces, but analytic outside the singular set

NN approximation of functions with isolated singularities

Exponential convergence of NN approximation for some elliptic PDEs
[M & Opschoor & Petersen & Schwab, Found. Comp. Math (2022)]

ReLU activation, $X = H^1(\Omega)$ in polygonal or polyhedral Ω, $\mathcal{A}$ solutions to elliptic equation with smooth coefficients. Then,

$$L(\varepsilon) = \mathcal{O}(|\log \varepsilon| \log |\log \varepsilon|), \qquad M(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d+1})$$

$\mathcal{A}$ contains weighted analytic functions with point and edge singularities, such as

- Solutions of elliptic PDEs with analytic coefficients and analytic rhs in
  - polygons
  - axiparallel polyhedra
- Eigenfunctions of elliptic eigenvalue problems with singular potentials
  - Linear & nonlinear Schrödinger
  - Hartree-Fock

NN approximation of functions with isolated singularities

Exponential convergence of NN approximation for some elliptic PDEs
[M & Opschoor & Petersen & Schwab, Found. Comp. Math (2022)]

ReLU activation, $X = H^1(\Omega)$ in polygonal or polyhedral Ω, $\mathcal{A}$ solutions to elliptic equation with smooth coefficients. Then,

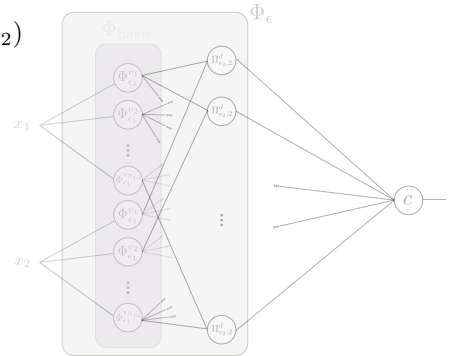
$$L(\varepsilon) = \mathcal{O}(|\log \varepsilon| \log |\log \varepsilon|), \qquad M(\varepsilon) = \mathcal{O}(|\log \varepsilon|^{2d+1})$$

$\mathcal{A}$ contains weighted analytic functions with point and edge singularities, such as

- Solutions of elliptic PDEs with analytic coefficients and analytic rhs in
  - polygons
  - axiparallel polyhedra
- Eigenfunctions of elliptic eigenvalue problems with singular potentials
  - Linear & nonlinear Schrödinger
  - Hartree-Fock

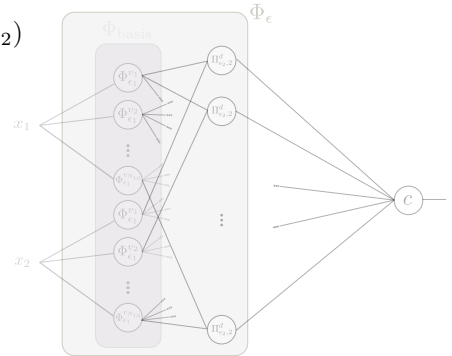
Sketch of the proof

- NN concatenation: $\mathbf{R}(\Phi_1 \odot \Phi_2) = \mathbf{R}(\Phi_1) \circ \mathbf{R}(\Phi_2)$
- NNs can approximate
  - multiplication
  - one dim. piecewise polynomials
- Construct a tensor product *hp* approximation
- Approximate one dim. polynomials with NNs
- Multiply with NNs to construct a basis
- Multiply by *hp* coefficients
- Error and size bounds from the components



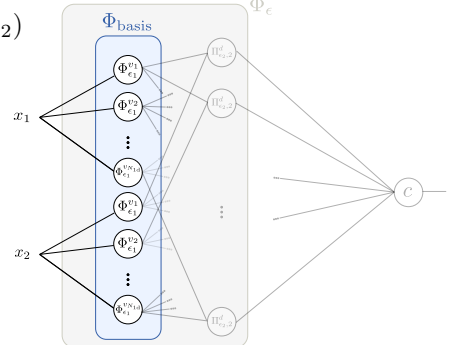
Sketch of the proof

- NN concatenation: $\mathbf{R}(\Phi_1 \odot \Phi_2) = \mathbf{R}(\Phi_1) \circ \mathbf{R}(\Phi_2)$
- NNs can approximate
  - multiplication
  - one dim. piecewise polynomials
- Construct a tensor product *hp* approximation
- Approximate one dim. polynomials with NNs
- Multiply with NNs to construct a basis
- Multiply by *hp* coefficients
- Error and size bounds from the components



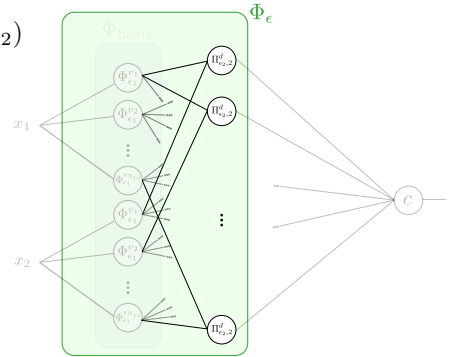
Sketch of the proof

- NN concatenation: $\mathrm{R}(\Phi_1 \odot \Phi_2) = \mathrm{R}(\Phi_1) \circ \mathrm{R}(\Phi_2)$
- NNs can approximate
  - multiplication
  - one dim. piecewise polynomials
- Construct a tensor product *hp* approximation
- Approximate one dim. polynomials with NNs
- Multiply with NNs to construct a basis
- Multiply by *hp* coefficients
- Error and size bounds from the components



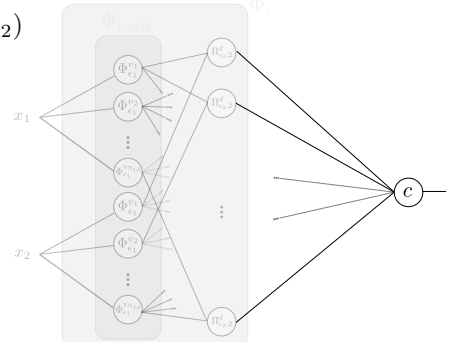
Sketch of the proof

- NN concatenation: $\mathbf{R}(\Phi_1 \odot \Phi_2) = \mathbf{R}(\Phi_1) \circ \mathbf{R}(\Phi_2)$
- NNs can approximate
  - multiplication
  - one dim. piecewise polynomials
- Construct a tensor product *hp* approximation
- Approximate one dim. polynomials with NNs
- Multiply with NNs to construct a basis
- Multiply by *hp* coefficients
- Error and size bounds from the components



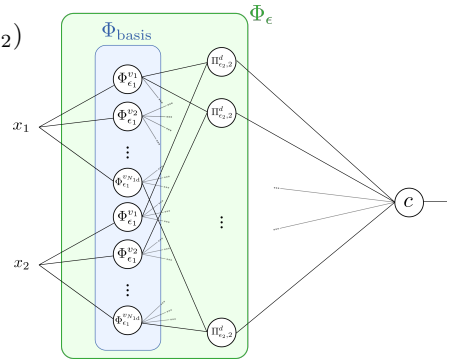
Sketch of the proof

- NN concatenation: $\mathbf{R}(\Phi_1 \odot \Phi_2) = \mathbf{R}(\Phi_1) \circ \mathbf{R}(\Phi_2)$
- NNs can approximate
  - multiplication
  - one dim. piecewise polynomials
- Construct a tensor product *hp* approximation
- Approximate one dim. polynomials with NNs
- Multiply with NNs to construct a basis
- Multiply by *hp* coefficients
- Error and size bounds from the components



Sketch of the proof

- NN concatenation: $\mathrm{R}(\Phi_1 \odot \Phi_2) = \mathrm{R}(\Phi_1) \circ \mathrm{R}(\Phi_2)$
- NNs can approximate
  - multiplication
  - one dim. piecewise polynomials
- Construct a tensor product *hp* approximation
- Approximate one dim. polynomials with NNs
- Multiply with NNs to construct a basis
- Multiply by *hp* coefficients
- Error and size bounds from the components



Operator Networks

Operator Networks

Universal approximation theorem for operators [Chen & Chen, IEEE Trans. Neural Netw. (1995)]

σ continuous non-polynomial function, X Banach space, $K_1 \Subset X$, $K_2 \Subset \mathbb{R}^d$, $V \subseteq C(K_1)$, $G : V \to C(K_2)$ nonlinear continuous operator. Then for any $\varepsilon > 0$,

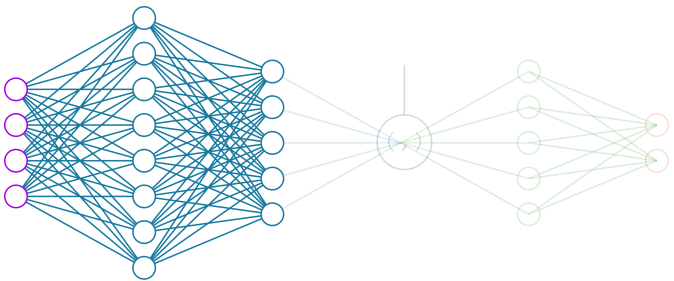
$$\left| G(a)(y) - \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k) \right| < \varepsilon \qquad \text{for all } a \in V,\ y \in K_2.$$

Operator Networks

Universal approximation theorem for operators [Chen & Chen, IEEE Trans. Neural Netw. (1995)]

σ continuous non-polynomial function, X Banach space, $K_1 \Subset X$, $K_2 \Subset \mathbb{R}^d$, $V \subseteq C(K_1)$, $G : V \to C(K_2)$ nonlinear continuous operator. Then for any $\varepsilon > 0$,

$$\left| G(a)(y) - \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k) \right| < \varepsilon \qquad \text{for all } a \in V,\ y \in K_2.$$

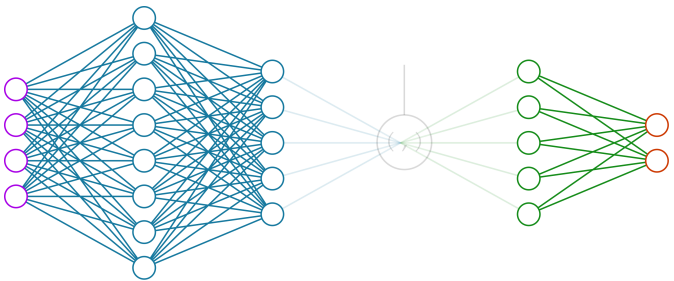


Operator Networks

Universal approximation theorem for operators [Chen & Chen, IEEE Trans. Neural Netw. (1995)]

σ continuous non-polynomial function, X Banach space, $K_1 \Subset X$, $K_2 \Subset \mathbb{R}^d$, $V \subset C(K_1)$, $G : V \to C(K_2)$ nonlinear continuous operator. Then for any $\varepsilon > 0$,

$$\left| G(a)(y) - \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k) \right| < \varepsilon \qquad \text{for all } a \in V,\ y \in K_2.$$

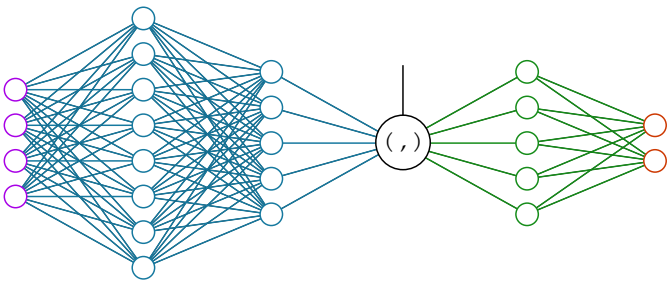


Operator Networks

Universal approximation theorem for operators [Chen & Chen, IEEE Trans. Neural Netw. (1995)]

σ continuous non-polynomial function, X Banach space, $K_1 \Subset X$, $K_2 \Subset \mathbb{R}^d$, $V \subseteq C(K_1)$, $G : V \to C(K_2)$ nonlinear continuous operator. Then for any $\varepsilon > 0$,

$$\left| G(a)(y) - \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k) \right| < \varepsilon \qquad \text{for all } a \in V,\ y \in K_2.$$



Deep Operator Networks

- More general structure [Lanthaler et al., Trans. Math. Appl. (2022)]:

$$\mathcal{X} \xrightarrow{\mathcal{E}} \mathbb{R}^n \xrightarrow{\mathcal{A}} \mathbb{R}^m \xrightarrow{\mathcal{R}} \mathcal{Y}$$

$$\mathcal{R} \circ \mathcal{A} \circ \mathcal{E} = \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k)$$

$$\mathcal{E} : a \mapsto \{a(x_j)\}_j \qquad \mathcal{A} : w \mapsto \mathrm{R}(\Phi_{\mathrm{app}})(w) \qquad \mathcal{R} : z \mapsto z^\top \mathrm{R}(\Phi_{\mathrm{reconst}})$$

- The two networks can be replaced by two deeper networks [Lu et at., Nat. Mach. Intell. (2021)]
- Encoding can also be, e.g., Fourier modes: Fourier Neural Operator [Li et al., ICLR (2021)]

Deep Operator Networks

- More general structure [Lanthaler et al., Trans. Math. Appl. (2022)]:

$$\mathcal{X} \xrightarrow{\mathcal{E}} \mathbb{R}^n \xrightarrow{\mathcal{A}} \mathbb{R}^m \xrightarrow{\mathcal{R}} \mathcal{Y}$$

$$\mathcal{R} \circ \mathcal{A} \circ \mathcal{E} = \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k)$$

$$\mathcal{E} : a \mapsto \{a(x_j)\}_j \qquad \mathcal{A} : w \mapsto \mathrm{R}(\Phi_{\mathrm{app}})(w) \qquad \mathcal{R} : z \mapsto z^\top \mathrm{R}(\Phi_{\mathrm{reconst}})$$

- The two networks can be replaced by two deeper networks [Lu et at., Nat. Mach. Intell. (2021)]
- Encoding can also be, e.g., Fourier modes: Fourier Neural Operator [Li et al., ICLR (2021)]

Deep Operator Networks

- More general structure [Lanthaler et al., Trans. Math. Appl. (2022)]:

$$\mathcal{X} \xrightarrow{\mathcal{E}} \mathbb{R}^n \xrightarrow{\mathcal{A}} \mathbb{R}^m \xrightarrow{\mathcal{R}} \mathcal{Y}$$

$$\mathcal{R} \circ \mathcal{A} \circ \mathcal{E} = \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k)$$

$$\mathcal{E} : a \mapsto \{a(x_j)\}_j \qquad \mathcal{A} : w \mapsto \mathrm{R}(\Phi_{\mathrm{app}})(w) \qquad \mathcal{R} : z \mapsto z^\top \mathrm{R}(\Phi_{\mathrm{reconst}})$$

- The two networks can be replaced by two deeper networks [Lu et at., Nat. Mach. Intell. (2021)]
- Encoding can also be, e.g., Fourier modes: Fourier Neural Operator [Li et al., ICLR (2021)]

Deep Operator Networks

- More general structure [Lanthaler et al., Trans. Math. Appl. (2022)]:

$$\mathcal{X} \xrightarrow{\quad\mathcal{E}\quad} \mathbb{R}^n \xrightarrow{\quad\mathcal{A}\quad} \mathbb{R}^m \xrightarrow{\quad\mathcal{R}\quad} \mathcal{Y}$$

$$\mathcal{R} \circ \mathcal{A} \circ \mathcal{E} = \sum_{k=1}^{m} \sum_{i=1}^{p} c_i^k \sigma \left(\sum_{j=1}^{n} \xi_{ij}^k a(x_j) + \theta_i^k \right) \sigma(w_k \cdot y + \zeta_k)$$

$$\mathcal{E} : a \mapsto \{a(x_j)\}_j \qquad \mathcal{A} : w \mapsto \mathrm{R}(\Phi_{\mathrm{app}})(w) \qquad \mathcal{R} : z \mapsto z^\top \mathrm{R}(\Phi_{\mathrm{reconst}})$$

- The two networks can be replaced by two deeper networks [Lu et at., Nat. Mach. Intell. (2021)]
- Encoding can also be, e.g., Fourier modes: Fourier Neural Operator [Li et al., ICLR (2021)]

A model elliptic PDE

$$-\nabla \cdot (a \nabla u^a) = f \quad \text{in } (\mathbb{R}/\mathbb{Z})^d$$

- Set of coefficients

$$\mathcal{D} = \left\{ \text{periodic } v : v(x) \geq a_{\min} > 0 \text{ and } \|v\|_{W^{k,\infty}(\Omega)} \leq A_{\mathcal{D}}^{k+1} k! \right\}$$

- Space of solutions ($Q = (0,1)^d$)

$$X = \left\{ v \in H^1(\Omega) : \int_Q v = 0 \right\},$$

- Fixed, analytic and periodic f
- Solution operator

$$S : \mathcal{D} \ni a \mapsto u^a \in X$$

A model elliptic PDE

$$-\nabla \cdot (a \nabla u^a) = f \quad \text{in } (\mathbb{R}/\mathbb{Z})^d$$

- Set of coefficients

$$\mathcal{D} = \left\{ \text{periodic } v : v(x) \geq a_{\min} > 0 \text{ and } \|v\|_{W^{k,\infty}(\Omega)} \leq A_{\mathcal{D}}^{k+1} k! \right\}$$

- Space of solutions ($Q = (0,1)^d$)

$$X = \left\{ v \in H^1(\Omega) : \int_Q v = 0 \right\},$$

- Fixed, analytic and periodic f
- Solution operator

$$S : \mathcal{D} \ni a \mapsto u^a \in X$$

A model elliptic PDE

$$-\nabla \cdot (a \nabla u^a) = f \quad \text{in } (\mathbb{R}/\mathbb{Z})^d$$

- Set of coefficients

$$\mathcal{D} = \left\{ \text{periodic } v : v(x) \geq a_{\min} > 0 \text{ and } \|v\|_{W^{k,\infty}(\Omega)} \leq A_{\mathcal{D}}^{k+1} k! \right\}$$

- Space of solutions ($Q = (0,1)^d$)

$$X = \left\{ v \in H^1(\Omega) : \int_Q v = 0 \right\},$$

- Fixed, analytic and periodic f
- Solution operator

$$S : \mathcal{D} \ni a \mapsto u^a \in X$$

A model elliptic PDE

$$-\nabla \cdot (a \nabla u^a) = f \quad \text{in } (\mathbb{R}/\mathbb{Z})^d$$

- Set of coefficients

$$\mathcal{D} = \left\{ \text{periodic } v : v(x) \geq a_{\min} > 0 \text{ and } \|v\|_{W^{k,\infty}(\Omega)} \leq A_{\mathcal{D}}^{k+1} k! \right\}$$

- Space of solutions ($Q = (0,1)^d$)

$$X = \left\{ v \in H^1(\Omega) : \int_Q v = 0 \right\},$$

- Fixed, analytic and periodic f
- Solution operator

$$S : \mathcal{D} \ni a \mapsto u^a \in X$$

A model elliptic PDE

$$-\nabla \cdot (a \nabla u^a) = f \quad \text{in } (\mathbb{R}/\mathbb{Z})^d$$

- Set of coefficients

$$\mathcal{D} = \left\{ \text{periodic } v : v(x) \geq a_{\min} > 0 \text{ and } \|v\|_{W^{k,\infty}(\Omega)} \leq A_{\mathcal{D}}^{k+1} k! \right\}$$

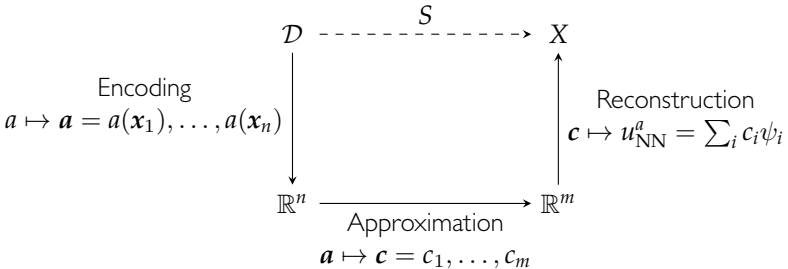
- Space of solutions ($Q = (0,1)^d$)

$$X = \left\{ v \in H^1(\Omega) : \int_Q v = 0 \right\},$$

- Fixed, analytic and periodic f
- Solution operator

$$S : \mathcal{D} \ni a \mapsto u^a \in X$$

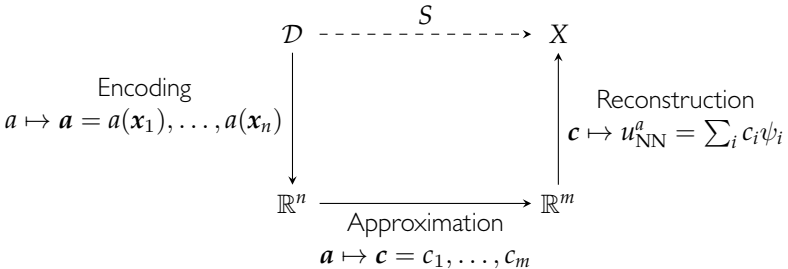
Operator networks again



$$\mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}} : a \mapsto \{a(x_j)\}_j \qquad \mathcal{A}_\Phi : w \mapsto \mathrm{R}(\Phi)(w) \qquad \mathcal{R}_\Phi : z \mapsto z^\top \mathrm{R}(\Phi)$$

$$a \mapsto u^a \quad \simeq \quad a \mapsto \left(\mathcal{R}_{\Phi^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}}\right)(a)$$

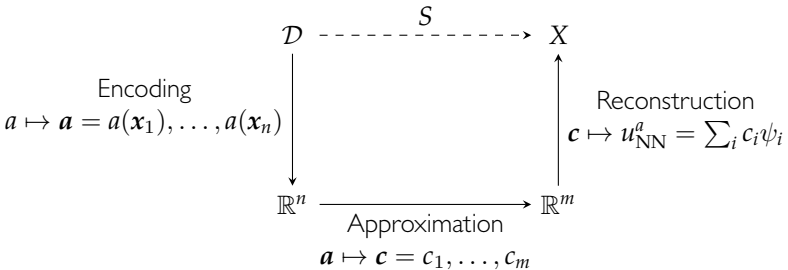
Operator networks again



$$\mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}} : a \mapsto \{a(x_j)\}_j \qquad \mathcal{A}_\Phi : w \mapsto \mathrm{R}(\Phi)(w) \qquad \mathcal{R}_\Phi : z \mapsto z^\top \mathrm{R}(\Phi)$$

$$a \mapsto u^a \quad \simeq \quad a \mapsto \left(\mathcal{R}_{\Phi^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}}\right)(a)$$

Operator networks again



$$\mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}} : a \mapsto \{a(x_j)\}_j \qquad \mathcal{A}_\Phi : w \mapsto \mathrm{R}(\Phi)(w) \qquad \mathcal{R}_\Phi : z \mapsto z^\top \mathrm{R}(\Phi)$$

$$a \mapsto u^a \quad \simeq \quad a \mapsto \left(\mathcal{R}_{\Phi^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}}\right)(a)$$

Operator network approximation

Approximation rates [M & Schwab, Submitted (2021)]

There exists $C > 0$ such that, for all $\varepsilon \in (0, 1)$, there exist

- $n_{\mathrm{b}}, n_{\mathrm{q}} \in \mathbb{N}$ and points $\boldsymbol{x}_{\mathrm{enc}} := \{\boldsymbol{x}_1, \dots \boldsymbol{x}_{n_{\mathrm{q}}}\} \subset \overline{Q}$,
- two NNs $\Phi_\varepsilon^{\mathrm{app}}$ and $\Phi_\varepsilon^{\mathrm{rec}}$ with $\mathrm{R}(\Phi_\varepsilon^{\mathrm{app}}) : \mathbb{R}^{n_{\mathrm{q}}} \to \mathbb{R}^{n_{\mathrm{b}}}$ and $\mathrm{R}(\Phi_\varepsilon^{\mathrm{rec}}) : Q \to \mathbb{R}^{n_{\mathrm{b}}}$,

such that

- $$\sup_{a \in \mathcal{D}} \|u^a - (\mathcal{R}_{\Phi_\varepsilon^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi_\varepsilon^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}})(a)\|_{H^1(\Omega)} \leq \varepsilon,$$
- $n_{\mathrm{b}}, n_{\mathrm{q}} \leq C(1 + |\log \varepsilon|^d)$,
- as $\varepsilon \downarrow 0$,

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon| \left(\log |\log \varepsilon|\right)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon|^{3d+2} \left(\log |\log \varepsilon|\right)^2\right),$$

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon| \left(\log |\log \varepsilon|\right)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon|^{d+1}\right).$$

Operator network approximation

Approximation rates [M & Schwab, Submitted (2021)]

There exists $C > 0$ such that, for all $\varepsilon \in (0,1)$, there exist

- $n_{\mathrm{b}}, n_{\mathrm{q}} \in \mathbb{N}$ and points $\boldsymbol{x}_{\mathrm{enc}} := \{\boldsymbol{x}_1, \dots \boldsymbol{x}_{n_{\mathrm{q}}}\} \subset \overline{Q}$,
- two NNs $\Phi_\varepsilon^{\mathrm{app}}$ and $\Phi_\varepsilon^{\mathrm{rec}}$ with $\mathrm{R}(\Phi_\varepsilon^{\mathrm{app}}) : \mathbb{R}^{n_{\mathrm{q}}} \to \mathbb{R}^{n_{\mathrm{b}}}$ and $\mathrm{R}(\Phi_\varepsilon^{\mathrm{rec}}) : Q \to \mathbb{R}^{n_{\mathrm{b}}}$,

such that

- $$\sup_{a \in \mathcal{D}} \|u^a - (\mathcal{R}_{\Phi_\varepsilon^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi_\varepsilon^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}})(a)\|_{H^1(\Omega)} \leq \varepsilon,$$
- $n_{\mathrm{b}}, n_{\mathrm{q}} \leq C(1 + |\log \varepsilon|^d)$,
- as $\varepsilon \downarrow 0$,

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon| \left(\log |\log \varepsilon|\right)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon|^{3d+2} \left(\log |\log \varepsilon|\right)^2\right),$$

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon| \left(\log |\log \varepsilon|\right)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon|^{d+1}\right).$$

Operator network approximation

Approximation rates [M & Schwab, Submitted (2021)]

There exists $C > 0$ such that, for all $\varepsilon \in (0,1)$, there exist

- $n_{\mathrm{b}}, n_{\mathrm{q}} \in \mathbb{N}$ and points $\boldsymbol{x}_{\mathrm{enc}} := \{\boldsymbol{x}_1, \dots \boldsymbol{x}_{n_{\mathrm{q}}}\} \subset \overline{Q}$,
- two NNs $\Phi_\varepsilon^{\mathrm{app}}$ and $\Phi_\varepsilon^{\mathrm{rec}}$ with $\mathrm{R}(\Phi_\varepsilon^{\mathrm{app}}) : \mathbb{R}^{n_{\mathrm{q}}} \to \mathbb{R}^{n_{\mathrm{b}}}$ and $\mathrm{R}(\Phi_\varepsilon^{\mathrm{rec}}) : Q \to \mathbb{R}^{n_{\mathrm{b}}}$,

such that

- $$\sup_{a \in \mathcal{D}} \|u^a - (\mathcal{R}_{\Phi_\varepsilon^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi_\varepsilon^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}})(a)\|_{H^1(\Omega)} \leq \varepsilon,$$
- $n_{\mathrm{b}}, n_{\mathrm{q}} \leq C(1 + |\log \varepsilon|^d)$,
- as $\varepsilon \downarrow 0$,

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon| \, (\log |\log \varepsilon|)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon|^{3d+2} \, (\log |\log \varepsilon|)^2\right),$$

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon| \, (\log |\log \varepsilon|)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon|^{d+1}\right).$$

Operator network approximation

Approximation rates [M & Schwab, Submitted (2021)]

There exists $C > 0$ such that, for all $\varepsilon \in (0,1)$, there exist

- $n_{\mathrm{b}}, n_{\mathrm{q}} \in \mathbb{N}$ and points $\boldsymbol{x}_{\mathrm{enc}} := \{\boldsymbol{x}_1, \dots \boldsymbol{x}_{n_{\mathrm{q}}}\} \subset \overline{Q}$,
- two NNs $\Phi_\varepsilon^{\mathrm{app}}$ and $\Phi_\varepsilon^{\mathrm{rec}}$ with $\mathrm{R}(\Phi_\varepsilon^{\mathrm{app}}) : \mathbb{R}^{n_{\mathrm{q}}} \to \mathbb{R}^{n_{\mathrm{b}}}$ and $\mathrm{R}(\Phi_\varepsilon^{\mathrm{rec}}) : Q \to \mathbb{R}^{n_{\mathrm{b}}}$,

such that

- $$\sup_{a \in \mathcal{D}} \|u^a - (\mathcal{R}_{\Phi_\varepsilon^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi_\varepsilon^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}})(a)\|_{H^1(\Omega)} \leq \varepsilon,$$
- $n_{\mathrm{b}}, n_{\mathrm{q}} \leq C(1 + |\log \varepsilon|^d)$,
- as $\varepsilon \downarrow 0$,

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon| \, (\log |\log \varepsilon|)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon|^{3d+2} \, (\log |\log \varepsilon|)^2\right),$$

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon| \, (\log |\log \varepsilon|)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon|^{d+1}\right).$$

Operator network approximation

Approximation rates [M & Schwab, Submitted (2021)]

There exists $C > 0$ such that, for all $\varepsilon \in (0,1)$, there exist

- $n_{\mathrm{b}}, n_{\mathrm{q}} \in \mathbb{N}$ and points $\boldsymbol{x}_{\mathrm{enc}} := \{\boldsymbol{x}_1, \dots \boldsymbol{x}_{n_{\mathrm{q}}}\} \subset \overline{Q}$,
- two NNs $\Phi_\varepsilon^{\mathrm{app}}$ and $\Phi_\varepsilon^{\mathrm{rec}}$ with $\mathrm{R}(\Phi_\varepsilon^{\mathrm{app}}) : \mathbb{R}^{n_{\mathrm{q}}} \to \mathbb{R}^{n_{\mathrm{b}}}$ and $\mathrm{R}(\Phi_\varepsilon^{\mathrm{rec}}) : Q \to \mathbb{R}^{n_{\mathrm{b}}}$,

such that

- $$\sup_{a \in \mathcal{D}} \|u^a - (\mathcal{R}_{\Phi_\varepsilon^{\mathrm{rec}}} \circ \mathcal{A}_{\Phi_\varepsilon^{\mathrm{app}}} \circ \mathcal{E}_{\boldsymbol{x}_{\mathrm{enc}}})(a)\|_{H^1(\Omega)} \leq \varepsilon,$$
- $n_{\mathrm{b}}, n_{\mathrm{q}} \leq C(1 + |\log \varepsilon|^d)$,
- as $\varepsilon \downarrow 0$,

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon| \left(\log |\log \varepsilon|\right)\right), \quad \mathrm{size}(\Phi_\varepsilon^{\mathrm{app}}) = \mathcal{O}\left(|\log \varepsilon|^{3d+2} \left(\log |\log \varepsilon|\right)^2\right),$$

$$\mathrm{depth}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon| \left(\log |\log \varepsilon|\right)\right), \qquad \mathrm{size}(\Phi_\varepsilon^{\mathrm{rec}}) = \mathcal{O}\left(|\log \varepsilon|^{d+1}\right).$$

Sketch of the proof

Idea: Approximate the fully discrete spectral-Galerkin solution

$$u^a_{n_\mathrm{b},n_\mathrm{q}}: \quad \sum_{k=1}^{n_\mathrm{q}} w_k a(x_k)(\nabla u^a_{n_\mathrm{b},n_\mathrm{q}})(x_k)(\nabla\varphi_i)(x_k) = \int_Q f\varphi_i, \qquad \forall i = 1,\dots,n_\mathrm{b},$$

- $\{w_k, x_k\}$ Gauss-Lobatto quadrature
- φ_i Legendre polynomials.

$$u^a_{n_\mathrm{b},n_\mathrm{q}} = \sum_{i=1}^{n_\mathrm{b}} c^a_i \varphi_i$$

Reconstruction network

$$[\mathbb{R}(\Phi^{\mathrm{rec}}_\varepsilon)]_i \simeq \varphi_i \quad \text{in } H^1(Q)$$

Sketch of the proof

Idea: Approximate the fully discrete spectral-Galerkin solution

$$u^a_{n_\mathrm{b},n_\mathrm{q}}: \quad \sum_{k=1}^{n_\mathrm{q}} w_k a(x_k)(\nabla u^a_{n_\mathrm{b},n_\mathrm{q}})(x_k)(\nabla\varphi_i)(x_k) = \int_Q f\varphi_i, \qquad \forall i = 1,\dots,n_\mathrm{b},$$

- $\{w_k, x_k\}$ Gauss-Lobatto quadrature
- φ_i Legendre polynomials.

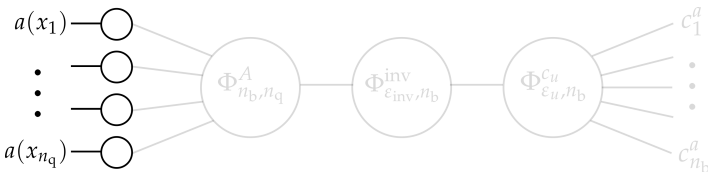
$$u^a_{n_\mathrm{b},n_\mathrm{q}} = \sum_{i=1}^{n_\mathrm{b}} c^a_i \varphi_i$$

Reconstruction network

$$[\mathrm{R}(\Phi^{\mathrm{rec}}_\varepsilon)]_i \simeq \varphi_i \quad \text{in } H^1(Q)$$

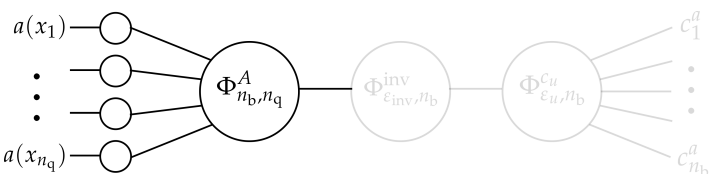
Approximation network

- $a(x_1), \dots a(x_{n_q})$
- $\mathrm{R}(\Phi^{A}_{n_b,n_q}) : a(x_1), \dots, a(x_{n_q}) \mapsto \sum_{k=1}^{n_q} w_k a(x_k)(\nabla\varphi_j)(x_k)(\nabla\varphi_i)(x_k) =: \left[A^a_{n_b,n_q}\right]_{ij}$
- $\mathrm{R}(\Phi^{\mathrm{inv}}_{\varepsilon_{\mathrm{inv}}}) : A^a_{n_b,n_q} \mapsto A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_b} \simeq \left(A^a_{n_b,n_q}\right)^{-1}$ [Kutyniok et al., Constr. Approx. (2021)]
- $\mathrm{R}(\Phi^{c_u}_{\varepsilon_u,n_b}) : A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_b} \mapsto A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_b} \left[\int_Q f\varphi_i\right]_i \simeq c^a$



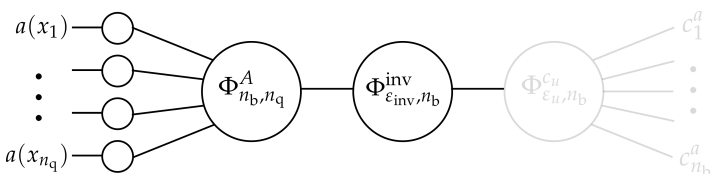
Approximation network

- $a(x_1), \dots a(x_{n_\mathrm{q}})$
- $\mathrm{R}(\Phi^A_{n_\mathrm{b},n_\mathrm{q}}) : a(x_1), \dots, a(x_{n_\mathrm{q}}) \mapsto \sum_{k=1}^{n_\mathrm{q}} w_k a(x_k)(\nabla\varphi_j)(x_k)(\nabla\varphi_i)(x_k) =: \left[A^a_{n_\mathrm{b},n_\mathrm{q}}\right]_{ij}$
- $\mathrm{R}(\Phi^{\mathrm{inv}}_{\varepsilon_{\mathrm{inv}}}) : A^a_{n_\mathrm{b},n_\mathrm{q}} \mapsto A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_\mathrm{b}} \simeq \left(A^a_{n_\mathrm{b},n_\mathrm{q}}\right)^{-1}$ [Kutyniok et al., Constr. Approx. (2021)]
- $\mathrm{R}(\Phi^{c_u}_{\varepsilon_u,n_\mathrm{b}}) : A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_\mathrm{b}} \mapsto A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_\mathrm{b}} \left[\int_Q f\varphi_i\right]_i \simeq c^a$



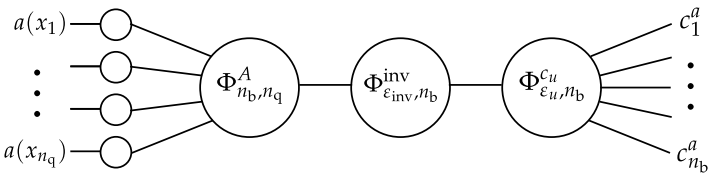
Approximation network

- $a(x_1), \dots a(x_{n_\mathrm{q}})$
- $\mathrm{R}(\Phi^A_{n_\mathrm{b},n_\mathrm{q}}) : a(x_1), \dots, a(x_{n_\mathrm{q}}) \mapsto \sum_{k=1}^{n_\mathrm{q}} w_k a(x_k)(\nabla\varphi_j)(x_k)(\nabla\varphi_i)(x_k) =: \left[A^a_{n_\mathrm{b},n_\mathrm{q}}\right]_{ij}$
- $\mathrm{R}(\Phi^{\mathrm{inv}}_{\varepsilon_{\mathrm{inv}}}) : A^a_{n_\mathrm{b},n_\mathrm{q}} \mapsto A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_\mathrm{b}} \simeq \left(A^a_{n_\mathrm{b},n_\mathrm{q}}\right)^{-1}$ [Kutyniok et al., Constr. Approx. (2021)]
- $\mathrm{R}(\Phi^{c_u}_{\varepsilon_u,n_\mathrm{b}}) : A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_\mathrm{b}} \mapsto A^a_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_\mathrm{b}} \left[\int_Q f\varphi_i\right]_i \simeq c^a$



Approximation network

- $a(x_1), \dots a(x_{n_q})$
- $\mathrm{R}(\Phi^{A}_{n_b,n_q}) : a(x_1), \dots, a(x_{n_q}) \mapsto \sum_{k=1}^{n_q} w_k a(x_k)(\nabla\varphi_j)(x_k)(\nabla\varphi_i)(x_k) =: \left[A^{a}_{n_b,n_q}\right]_{ij}$
- $\mathrm{R}(\Phi^{\mathrm{inv}}_{\varepsilon_{\mathrm{inv}}}) : A^{a}_{n_b,n_q} \mapsto A^{a}_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_b} \simeq \left(A^{a}_{n_b,n_q}\right)^{-1}$ [Kutyniok et al., Constr. Approx. (2021)]
- $\mathrm{R}(\Phi^{c_u}_{\varepsilon_u,n_b}) : A^{a}_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_b} \mapsto A^{a}_{\mathrm{inv};\varepsilon_{\mathrm{inv}},n_b} \left[\int_Q f\varphi_i\right]_i \simeq c^a$



Extensions

- Parametric PDEs

$$-\nabla \cdot (a(\boldsymbol{y})\nabla u) = f \qquad \boldsymbol{y} \in \mathbb{R}^{n_{\mathrm{P}}} \text{ parameter}$$

- Anistropic diffusion-reaction

$$-\nabla \cdot (\mathbf{A}\nabla u) + cu = f$$

- Linear elasticity

$$-\nabla \cdot (\boldsymbol{\sigma}[\boldsymbol{u}]) = \boldsymbol{f}$$

Some open questions & perspectives

- Operator networks for problems with boundary
- High-dimensional problems
- General activation functions
- Training (and computations)

Thank you for your attention