Error bounds for physics-informed (and) operator learning for PDEs

Tim De Ryck

ETH Zürich
Joint work with Prof. Siddhartha Mishra and Roberto Molinaro

FAU DCN-AvH Erlangen
20 April 2023

Neural networks

A **feedforward (artificial) neural network** of depth L is a map of the form

$$u_\theta : \mathbb{R}^{n_0} \to \mathbb{R}^{n_L} : x \mapsto (\mathcal{A}^L \circ \rho^{L-1} \circ \cdots \circ \rho^1 \circ \mathcal{A}^1)(x)$$

where

- $\mathcal{A}^\ell : \mathbb{R}^{n_{\ell-1}} \to \mathbb{R}^{n_\ell}$ are affine linear maps,
- ρ^ℓ are activation functions,
- θ are all parameters (weights and biases).

A feedforward NN is called a **tanh NN** when

- $\tanh : \mathbb{R} \to \mathbb{R} : x \mapsto \frac{e^x - e^{-x}}{e^x + e^{-x}}$,
- $\rho^\ell(x) = (\tanh(x_1), \ldots, \tanh(x_{n_\ell}))$ for $\ell = 1, 2, \ldots, L-1$,

Network has depth L, width $\max_\ell n_\ell$ and $\sum_\ell n_\ell$ neurons

Setting

Given $T > 0$ and $D \subset \mathbb{R}^d$ compact, consider PDE (with parameter a):

$$\mathcal{L}_a(u)(t,x) = 0 \quad \text{and} \quad u(x,0) = u_0 \quad (\mathcal{B}u = 0) \qquad \forall (t,x) \in [0,T] \times D,$$

e.g. heat equation $\mathcal{L}_a = \partial_t - a \cdot \Delta_x$

We are interested in:

Task 1: approximate function u for fixed u_0 and a

To do: find NN u_θ such that $u \approx u_\theta$

Task 2: approximate operator $\mathcal{G} : \mathcal{X} \to \mathcal{Y} : v \mapsto u$, where $v \in \{u_0, a\}$

To do: find NN operator $\mathcal{G}_\theta$ such that $\mathcal{G} \approx \mathcal{G}_\theta$

Why neural networks?

Task 1: supervised learning

Goal: find NN u_θ such that $\|u - u_\theta\|_{L^2}$ is small

Supervised learning consists of

- select (grid) points y_i and use PDE solver to approximate $u(y_i)$
- training set $\mathcal{S} = \{(y_1, u(y_1)), \ldots (y_N, u(y_N))\} \subset D \times u(D)$,
- approach: minimize $\frac{1}{N}\sum_{n=1}^{N} \|u(y_n) - u_\theta(y_n)\|^2$,
- problem: generating training data can be expensive,

Task 1: physics-informed learning

Goal: find NN u_θ such that $\|u - u_\theta\|_{L^2}$ is small

Problem: generating training data can be expensive

Recall $\mathcal{L}u = \mathcal{B}u = 0 \Rightarrow$ idea: minimize residuals $\|\mathcal{L}(u_\theta)\|_{L^2} + \lambda\|\mathcal{B}(u_\theta)\|_{L^2}$

Physics informed (unsupervised) learning consists of

- e.g. [Lagaris et al., 2000; Raissi et al., 2019],
- select (grid) points $x_i \in D$ and $y_i \in \partial D$
- training sets $\{x_1, \dots x_N\} \subset D$ and $\{y_1, \dots y_M\} \subset \partial D$ are free
  $\rightarrow$ no data generation necessary,
- approach: minimize $\frac{1}{N}\sum_{n=1}^{N} \|\mathcal{L}(u_\theta)(x_n)\| + \frac{\lambda}{M}\sum_{m=1}^{M} \|\mathcal{B}(u_\theta)(y_m)\|$.

Task 1: physics-informed learning

Goal: find NN u_θ such that $\|u - u_\theta\|_{L^2}$ is small

$\mathcal{D}u = \mathcal{B}u = 0 \Rightarrow$ idea: minimize PINN residual $\|\mathcal{L}(u_\theta)\|_{L^2} + \lambda\|\mathcal{B}(u_\theta)\|_{L^2}$

Physics informed (unsupervised) learning consists of

- training sets $\{x_1, \ldots x_N\} \subset D$ and $\{y_1, \ldots y_M\} \subset D$,
- approach: minimize $\frac{1}{N}\sum_{n=1}^{N} \|\mathcal{L}(u_\theta)(x_n)\| + \frac{\lambda}{M}\sum_{m=1}^{M} \|\mathcal{B}(u_\theta)(y_m)\|$.

Questions

1. **Existence**: Is there u_θ such that $\|\mathcal{D}u_\theta\|_{L^2} + \lambda\|\mathcal{B}u_\theta\|_{L^2}$ is small? If yes, what is the size of u_θ?

Task 1: physics-informed learning

Goal: find NN u_θ such that $\|u - u_\theta\|_{L^2}$ is small

$\mathcal{D}u = \mathcal{B}u = 0 \Rightarrow$ idea: minimize PINN residual $\|\mathcal{L}(u_\theta)\|_{L^2} + \lambda \|\mathcal{B}(u_\theta)\|_{L^2}$

Physics informed (unsupervised) learning consists of

- training sets $\{x_1, \ldots x_N\} \subset D$ and $\{y_1, \ldots y_M\} \subset D$,
- approach: minimize $\frac{1}{N}\sum_{n=1}^{N} \|\mathcal{L}(u_\theta)(x_n)\| + \frac{\lambda}{M}\sum_{m=1}^{M} \|\mathcal{B}(u_\theta)(y_m)\|$.

Questions

1. **Existence**: Is there u_θ such that $\|\mathcal{D}u_\theta\|_{L^2} + \lambda\|\mathcal{B}u_\theta\|_{L^2}$ is small? If yes, what is the size of u_θ?
2. **Stability**: If $\|\mathcal{D}u_\theta\|_{L^2} + \lambda\|\mathcal{B}u_\theta\|_{L^2}$ is small, will $\|u - u_\theta\|_{L^2}$ be small as well?

Task 1: physics-informed learning

Goal: find NN u_θ such that $\|u - u_\theta\|_{L^2}$ is small

$\mathcal{D}u = \mathcal{B}u = 0 \Rightarrow$ idea: minimize PINN residual $\|\mathcal{L}(u_\theta)\|_{L^2} + \lambda \|\mathcal{B}(u_\theta)\|_{L^2}$

Physics informed (unsupervised) learning consists of

- training sets $\{x_1, \dots x_N\} \subset D$ and $\{y_1, \dots y_M\} \subset D$,
- approach: minimize $\frac{1}{N}\sum_{n=1}^{N} \|\mathcal{L}(u_\theta)(x_n)\| + \frac{\lambda}{M}\sum_{m=1}^{M} \|\mathcal{B}(u_\theta)(y_m)\|$.

Questions

1. **Existence**: Is there u_θ such that $\|\mathcal{D}u_\theta\|_{L^2} + \lambda\|\mathcal{B}u_\theta\|_{L^2}$ is small? If yes, what is the size of u_θ?
2. **Stability**: If $\|\mathcal{D}u_\theta\|_{L^2} + \lambda\|\mathcal{B}u_\theta\|_{L^2}$ is small, will $\|u - u_\theta\|_{L^2}$ be small as well?
3. **Generalization**: Does small training error imply small generalization error?

Approximation PDEs with NNs

What needs to be minimized?

Supervised learning: training data needed

Physics-informed learning: no training data needed

- motivation: $\mathcal{L}(u) = 0$ so find u_θ with $\mathcal{L}(u_\theta) \approx 0$

Operator learning: multiple architectures, e.g.

- Deep Operator Network (DeepONet)
- Fourier Neural Operator (FNO)

$\mathcal{J}(\theta)$	supervised learning	physics-informed learning
Task 1 (function)	$\|u - u_\theta\|_{L^2(D\times[0,T])}$	$\|\mathcal{L}(u_\theta)\|_{L^2(D\times[0,T])}$
Task 2 (operator)	$\|\mathcal{G} - \mathcal{G}_\theta\|_{L^2(\mathcal{X}\times D\times[0,T])}$	$\|\mathcal{L}(\mathcal{G}_\theta)\|_{L^2(\mathcal{X}\times D\times[0,T])}$

Neural network approximation theory

Focus on (1): **How large should my NN be such that $\mathcal{J}(\theta) < \varepsilon$?**

- Find bounds on width, depth, weights ...
- **Curse of dimensionality (CoD)** overcome if size is $\mathcal{O}(\mathrm{pol}(d\varepsilon^{-1}))$

Are there any theoretical results?

Neural network approximation theory

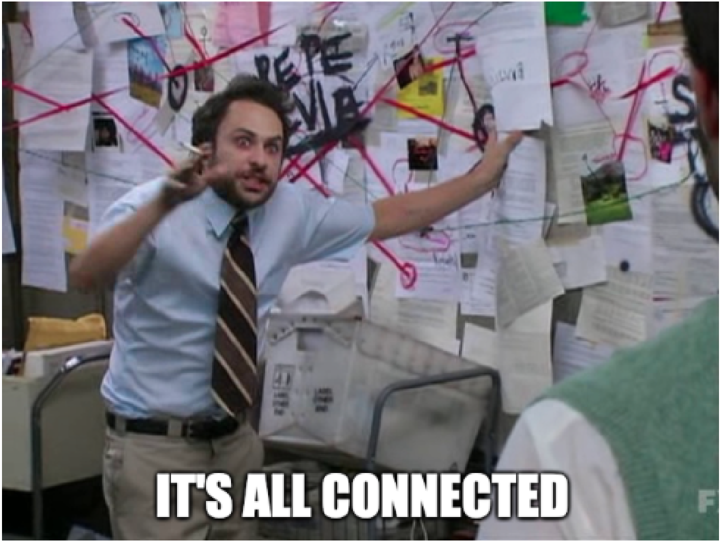
Focus on (1): **How large should my NN be such that $\mathcal{J}(\theta) < \varepsilon$?**

- Find bounds on width, depth, weights ...
- **Curse of dimensionality (CoD)** overcome if size is $\mathcal{O}(\mathrm{pol}(d\varepsilon^{-1}))$

Are there any theoretical results?

- Vanilla NN (supervised learning): yes, fairly easy, case-by-case
    - emulate FEM / FDS / FVM / MC / iterative method with NNs
- Physics-informed NN:
    - Low-dimensional and regular: yes, general bounds
    - High-dimensional: more difficult, case-by-case
- Operator learning: limited results, case-by-case
- Physics-informed operator learning: no results

Main message of paper



Generic bounds

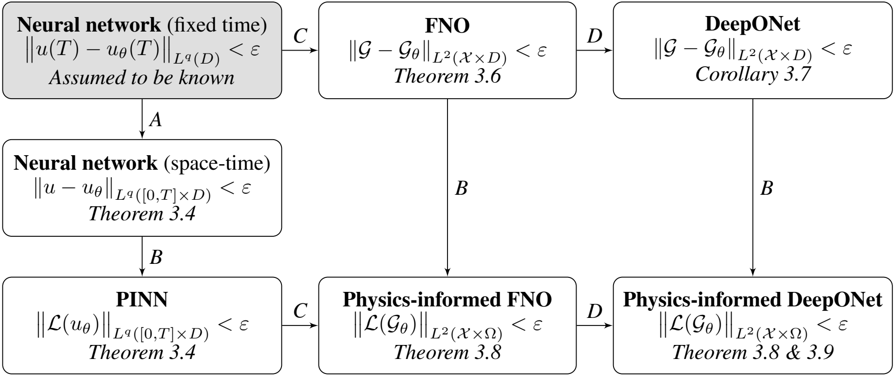


Figure: Visualization of how different types of error estimates can be obtained from one another. The letters reflect the techniques used in the proofs.

Assumptions

Assumption (1 - Fixed-time NN)

For every t, v there is a NN $\mathcal{U}^{\varepsilon}(v, t)$ that approximates u to accuracy ε. Moreover, there is an upper bound on the derivatives of the NN.

Assumption (2 - Domain)

Domain D is rectangular or smooth. (simplified)

Assumption (3 - Derivatives)

$$\|\mathcal{L}(u_\theta)\|_{L^q([0,T]\times D)} \leq C \cdot \sum_{\substack{(k',\boldsymbol{\alpha})\in\mathbb{N}_0^{d+1} \\ k'\leq k, \|\boldsymbol{\alpha}\|_1\leq \ell}} \left\| D^{(k',\boldsymbol{\alpha})}(u - u_\theta) \right\|_{L^q([0,T]\times D)}$$

Assumptions

Assumption (1 - Fixed-time NN)

For every t, v there is a NN $\mathcal{U}^{\varepsilon}(v, t)$ that approximates u to accuracy ε. Moreover, there is an upper bound on the derivatives of the NN.

Assumption (2 - Domain)

Domain D is rectangular or smooth. (simplified)

Assumption (3 - Derivatives)

$$\|\mathcal{L}(u_\theta)\|_{L^q([0,T]\times D)} \leq C \cdot \sum_{\substack{(k',\boldsymbol{\alpha})\in\mathbb{N}_0^{d+1} \\ k'\leq k, \|\boldsymbol{\alpha}\|_1\leq \ell}} \left\| D^{(k',\boldsymbol{\alpha})}(u - u_\theta)\right\|_{L^q([0,T]\times D)}$$

e.g. for heat equation $\mathcal{L}(u_\theta) = \partial_t u_\theta - \partial_x^2 u_\theta$ we get:

$$\|\mathcal{L}(u_\theta)\|_{L^q} = \|\mathcal{L}(u_\theta) - \mathcal{L}(u)\|_{L^q} \leq \|\partial_t(u - u_\theta)\|_{L^q} + \left\|\partial_x^2(u - u_\theta)\right\|_{L^q}$$

Assumptions

Assumption (1 - Fixed-time NN)

For every t, u_0 there is a NN $\mathcal{U}^\varepsilon(u_0, t)$ that approximates u to accuracy ε. Moreover, there is an upper bound on the derivatives of the NN.

Assumption (2 - Domain)

Domain D is rectangular or smooth. (simplified)

Assumption (3 - Derivatives)

$$\|\mathcal{L}(u_\theta)\|_{L^q([0,T]\times D)} \leq C \cdot \sum_{\substack{(k',\boldsymbol{\alpha})\in\mathbb{N}_0^{d+1} \\ k'\leq k, \|\boldsymbol{\alpha}\|_1\leq \ell}} \left\| D^{(k',\boldsymbol{\alpha})}(u - u_\theta) \right\|_{L^q([0,T]\times D)}$$

Assumption (4 - Stability)

$$\left\|\mathcal{U}^\varepsilon(u_0, T) - \mathcal{U}^\varepsilon(u_0', T)\right\|_{L^2} \leq C_{\text{stab}}^\varepsilon \left\| u_0 - u_0' \right\|_{L^p}. \quad (1)$$

Results

Informal summary of our results:

Theorem

If there is an approximation result for X that satisfies Y, then we can prove an approximation result for Z, with X + Y ⇒ Z being,

- *Fixed-time NN + Ass. 1 ⇒ space-time NN*
- *Fixed-time NN + Ass. 1,2,3 ⇒ PINN*
- *Operator learning + Ass. 2,4 ⇒ physics-informed operator learning*
- *Fixed-time NN + Ass. 1,3,4 ⇒ (physics-informed) operator learning*

Moreover,

- *FNO ⇒ DeepONet [Lanthaler, Mishra, Karniadakis; 2022]*
- *Smooth operator + Ass. 4 ⇒ operator learning.*

Example

Theorem

Let $r, s \in \mathbb{N}$, *let* $u \in C^{(s,r)}([0,T] \times D)$ *be the solution of the PDE and let Assumption 1 be satisfied. There exists a constant* $C(s,r) > 0$ *such that for every* $M \in \mathbb{N}$ *and* $\varepsilon, h > 0$ *there exists a tanh neural network* $u_\theta : [0,T] \times D \to \mathbb{R}$ *for which it holds that,*

$$\|u_\theta - u\|_{L^q([0,T]\times D)} \leq C(\|u\|_{C^{(s,0)}} M^{-s} + \varepsilon). \tag{2}$$

and if additionally Assumption 2 and Assumption 3 hold then,

$$\|\mathcal{L}(u_\theta)\|_{L^2([0,T]\times D)} + \|u_\theta - u\|_{L^2(\partial([0,T]\times D))} \leq C \cdot \mathrm{poly}(d) \cdot \ln^k(M)(\|u\|_{C^{(s,\ell)}} M^{k-s} + M^{2k}(\varepsilon h^{-\ell} + C^B_{\varepsilon,\ell} h^{r-\ell})). \tag{3}$$

Moreover, $\mathrm{depth}(u_\theta) \leq C \cdot \mathrm{depth}(\mathcal{U}^\varepsilon)$ *and* $\mathrm{width}(u_\theta) \leq CM \cdot \mathrm{width}(\mathcal{U}^\varepsilon)$.

Applications

Linear Kolmogorov PDEs

- PINNs overcome CoD

Nonlinear parabolic PDEs, e.g. Allen-Cahn equation

- PINNs overcome CoD
- (Physics-informed) operator learning: dimension-independent convergence rate for smooth functions

First results for **physics-informed operator learning**

- Linear operator: general result
- Nonlinear operator: case-by-case

Are we done now?

Restriction: results only valid for **smooth PDE solutions**

PDE might only have weak solution
$\Rightarrow$ PINN loss $\|\mathcal{L}(u)\|_{L^2}$ can not be evaluated,

Example: **scalar conservation laws**: $\mathcal{L}(u) := \partial_t u + \partial_x f(u) = 0$

- **Weak solutions**: u such that for all test functions φ we have $\int_{\mathbb{R}_+} \int_{\mathbb{R}} (u\varphi_t + f(u)\varphi_x)\, dxdt + \int_{\mathbb{R}} u_0(x)\varphi(x,0)dx = 0$
- **Entropy solutions**: weak solution + for every entropy pair (η, q) must hold that $\partial_t \eta(u) + \partial_x q(u) \leq 0$ (dist.)

Weak PINNs

We can restrict to *Kruzkhov entropy pairs*: $\partial_t \eta(u) + \partial_x q(u) \leq 0$ with,

- $\eta(u) = |u - c|$ and,
- $q(u) = Q[u; c] = \mathrm{sign}(u - c)(f(u) - f(c)$ for all $c \in \mathbb{R}$.

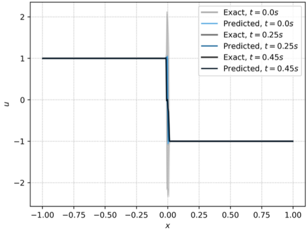
Result: u is entropy solution if $\forall \varphi$ (test functions) and $\forall c \in \mathbb{R}$,

$$0 \geq \mathcal{R}(u, \varphi, c) := - \int_D \int_{[0,T]} (|u(x,t) - c| \partial_t \varphi(x,t) + Q[u(x,t); c] \partial_x \varphi(x,t))$$

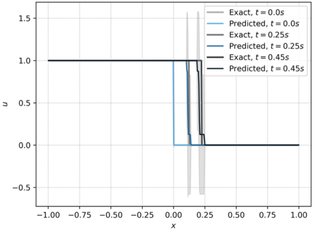
Idea of **weak PINNs**: find NN u_θ that minimizes

$$\mathcal{J}(u_\theta; \Phi) := \sup_{c \in \mathbb{R}, \varphi \in \Phi} \mathcal{R}(u_\theta, \varphi, c) + \|u_\theta(t=0) - u_0\|_{L^2} + (\text{BC term})$$

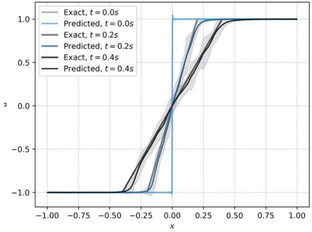
Weak PINNs experiments



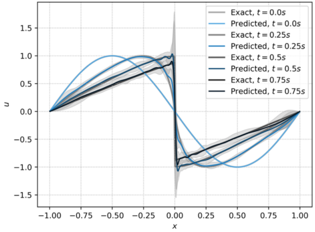
(A) Standing Shock Solution , $\mathcal{E}_r^T(\theta_{\mathcal{S}}^*) = 0.01$



(B) Moving Shock Solution, $\mathcal{E}_r^T(\theta_{\mathcal{S}}^*) = 0.019$



(C) Rarefaction Wave, $\mathcal{E}_r^T(\theta_{\mathcal{S}}^*) = 0.022$



(D) Initial Sine Wave, $\mathcal{E}_r^T(\theta_{\mathcal{S}}^*) = 0.057$

Theory for weak PINNs

We answer our three initial central questions for physics-informed learning:

1. **Existence**: for every $\varepsilon > 0$ there exists a NN u_θ such that $\mathcal{J}(u_\theta) < \varepsilon$

2. **Stability**: for every $\varepsilon > 0$ and a set Φ_ε (larger for smaller ε) it holds,

$$\|u_\theta(T) - u(T)\|_{L^1} \lesssim \varepsilon + \mathcal{J}(u_\theta; \Phi_\varepsilon)$$

3. **Generalization**: for M collocation points we have,

$$\|u_\theta(T) - u(T)\|_{L^1} \lesssim \varepsilon + (\text{training error}) + \frac{1}{\sqrt{M}}$$

Summary

Main points:

- Generic error bounds when PDE solutions are *sufficiently regular*
- Hyperbolic conservation laws: wPINNs recover *entropy solutions*

References:

- De Ryck, T. and Mishra, S. *Generic bounds on the approximation error for physics-informed (and) operator learning.* NeurIPS, 2022.
- De Ryck, T., Mishra, S. and Molinaro, R. *Weak physics-informed neural networks for approximating entropy solution of hyperbolic conservation laws.* Preprint, 2022.