Operator learning for history-dependent and multiscale problems

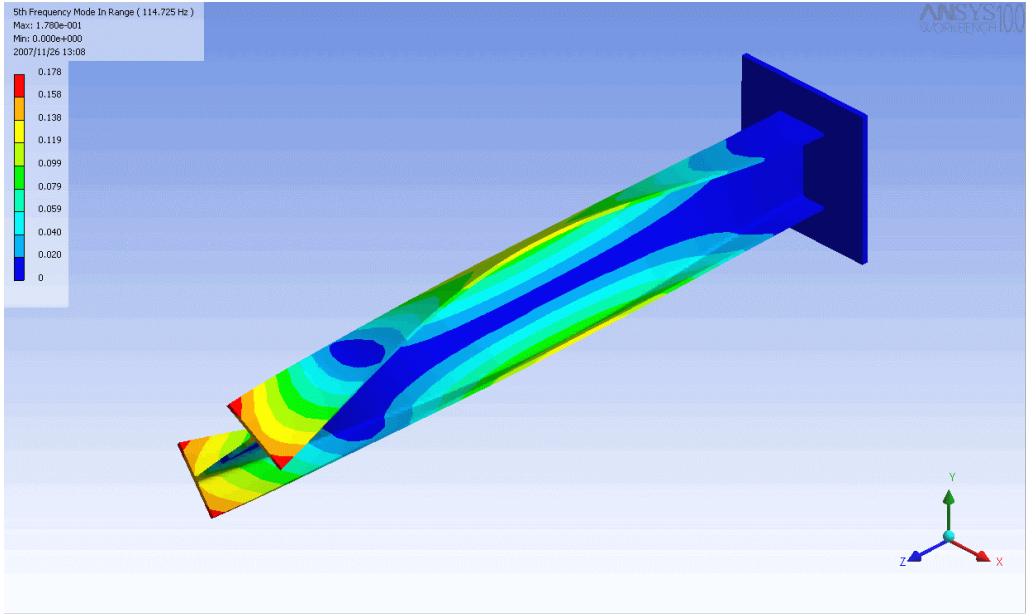
Margaret Trautner

Data Assimilation and Inverse Problems for Digital Twins

October 8, 2025

Caltech

Motivation via example: predicting material dynamics



Caltech

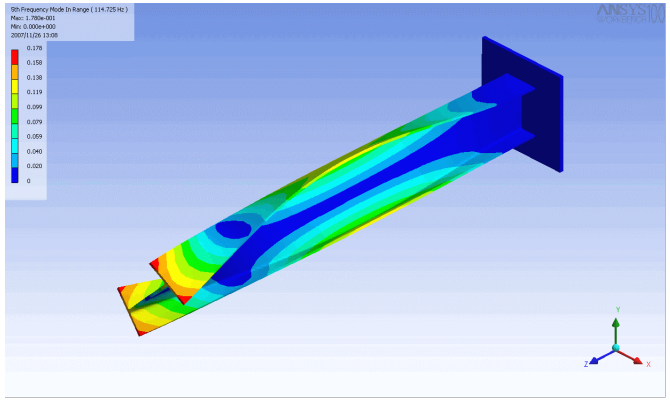
Constitutive laws

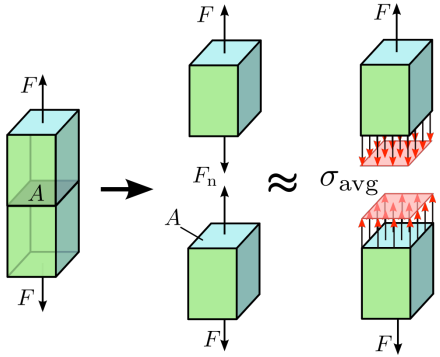
Equations for material dynamics:

1. $F = m \cdot a$
2. Initial condition
3. Boundary condition
4. Constitutive law: relates material displacement to force (material strain to material stress)

Stress (normal): force per area

Strain: gradient of displacement





Caltech

Simple constitutive law: linear elasticity

Mass (density) Acceleration Force

Force balance $$\rho \partial_t^2 u(x,t) = \nabla \cdot \sigma(x,t) + f(x,t)$$

Constitutive law (linear elasticity) $$\sigma(x,t) = A \nabla_x u(x,t)$$

Zero initial conditions

Zero Dirichlet boundary conditions

Caltech

Multiscale constitutive law

Displacement Stress External forcing

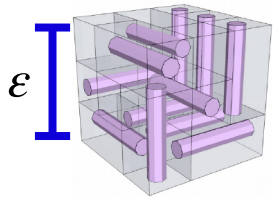
Force balance $$\rho \partial_t^2 u_\varepsilon(x,t) = \nabla_x \cdot \sigma_\varepsilon(x,t) + f(x,t),$$

Constitutive law $$\sigma_\varepsilon(x,t) = \Psi_\varepsilon^\dagger\big(\{\nabla_x u_\varepsilon(x,s)\}_{s\in[0,t]}, M\big)(t)$$

Strain trajectory

Zero initial conditions

Zero Dirichlet boundary conditions



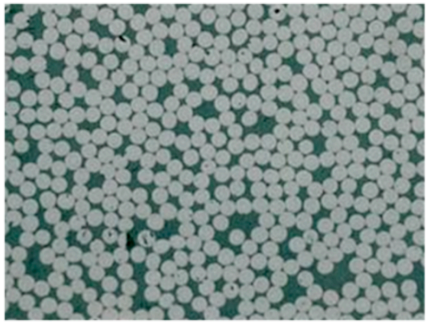
ε indicates dependence on small scale variable $y = \dfrac{x}{\varepsilon}$

Caltech

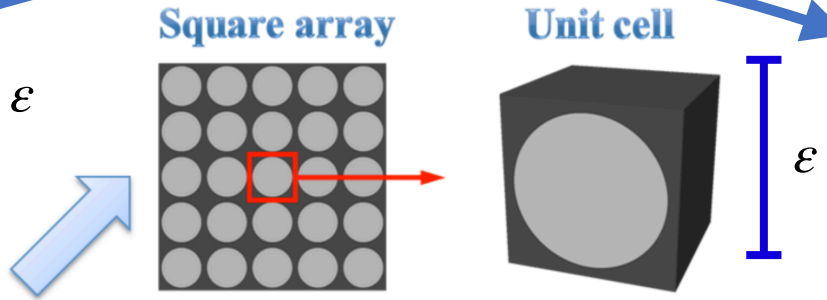
Zhang, Mingzhong, 2013.

Approaching multiscale problems

1. Assume periodicity in ε



Cross-sectional view of continuous fiber reinforced composites



2. Use known physics to solve PDE on unit cell

3. Use resultant stress on all repeated cells to find displacement at next time iteration

Caltech

Homogenized equations

Displacement

Stress

External forcing

Force balance $$\rho \partial_t^2 u_0(x,t) = \nabla_x \cdot \sigma_0(x,t) + f(x,t),$$

Constitutive law $$\sigma_0(x,t) = \Psi_0^\dagger\big(\{\nabla_x u_0(x,s)\}_{s\in[0,t]}; M\big)(t)$$

Strain trajectory

Zero initial conditions

Zero Dirichlet boundary conditions

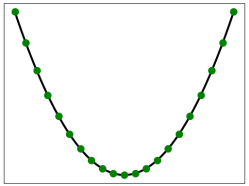
- Assume microstructure ε periodic
- Seek *homogenized constitutive law* $\Psi_0^\dagger$ such that $u_\varepsilon \to u_0$ as $\varepsilon \to 0$
- $\Psi_0^\dagger$: homogenized strain $\mapsto$ homogenized stress
- Homogenized constitutive law maps between *functions* in time

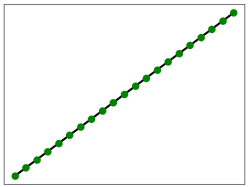
Caltech

Operator learning

- Classical machine learning methods map between finite-dimensional spaces
- PDEs map between function spaces
- Operator learning approximates *infinite-dimensional operators* using data

Example: Learn derivative map $\frac{d}{dx} : C^1([0,1]) \mapsto C([0,1])$



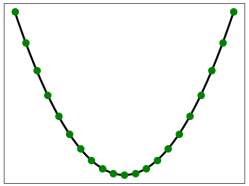


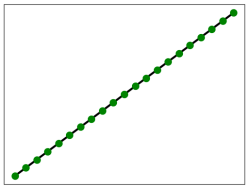
	Learning in Finite Dimensions	Learning in Infinite Dimensions
Mapping between	$\mathbb{R}^{21} \mapsto \mathbb{R}^{21}$	$C^1([0,1]) \mapsto C([0,1])$
Discretization	Hardwired into model architecture	Affects learned model parameters only indirectly
Evaluation at new output points	Can only be interpolated from model output points	Model itself can be evaluated at new points

When is operator learning useful?

- Classical methods work well to solve PDEs
- Operator learning only potentially useful if:

1. Data generation/training cost can be amortized
2. Equations are unknown





	Learning in Finite Dimensions	Learning in Infinite Dimensions
Mapping between	$\mathbb{R}^{21} \mapsto \mathbb{R}^{21}$	$C^1([0,1]) \mapsto C([0,1])$
Discretization	Hardwired into model architecture	Affects learned model parameters only indirectly
Evaluation at new output points	Can only be interpolated from model output points	Model itself can be evaluated at new points

Some existing work on operator learning

I. Architectures

- Fourier Neural Operators (FNO) (Li et al. 2021), DeepONet (Lu et al. 2021), GNO (Li et al. 2020), PCA-Net (Liu et al. 2022), ANO (Lanthaler et al. 2023), etc.

II. Analysis

- Universal approximation: FNO (Kovachki et al. 2021), DeepONet (Lu et al. 2021), etc. UA in finite dimensions goes back to Cybenko (1989).
- Error bounds with respect to model size: FNO (Kovachki et al. 2021), DeepONet (Lanthaler et al. 2021), DON with Lipschitz/Holder maps (Schwab et al. 2023) etc. Foundational work in finite dimensions by Yarotsky (2017).

III. Application setting of constitutive modeling in solid mechanics

- PCA-Net approach to multiscale setting (Liu et al. 2022), deep learning approach to constitutive models (Huang et al. 2020) (Mozaffar et al. 2019), structure-preserving deep learning approach to constitutive models (As'ad et al. 2022), etc.

Outline: Operator learning for history-dependent and multiscale problems

I. Recurrent neural operator for causal history-dependent models

II. Learning for discontinuous inputs (materials) with FNO

III. Comparison of parameter-to-observable operator learning for multiscale elasticity via Fourier Neural Mapping (FNM)

IV. Sampling convergence rate theory-to-practice gap

Caltech

History-dependent constitutive laws

- For some materials, stress depend on entire strain history

- Kelvin-Voigt viscoelastic material

$$\sigma_\varepsilon(t) = E_\varepsilon \text{strain}_\varepsilon(t) + \nu_\varepsilon \partial_t \text{strain}_\varepsilon(t)$$

Homogenized:

$$\sigma(t) = E' \text{strain}(t) + \nu' \partial_t \text{strain}(t) - \int_0^t \kappa(t-\tau) \text{strain}(\tau)\, \mathrm{d}\tau$$

Caltech

Markovian model to capture history dependence

Recurrent neural operator (RNO) architecture

$$\sigma(t) = F(\text{strain}(t), \partial_t \text{strain}(t), \xi(t))$$

$$\dot{\xi}(t) = G(\text{strain}(t), \xi(t))$$

$$\xi(0) = 0$$

- $\xi \in \mathbb{R}^L$ are hidden variables
- F and G are neural networks
- ξ updated with any time integration scheme

- ξ captures memory
- Discretization invariant

Caltech

RNO model and 1d viscoelasticity

- For piecewise-constant viscoelastic materials, homogenized constitutive law takes the form

$$\sigma(t) = E'\text{strain}(t) + \nu'\partial_t\text{strain}(t) - \langle \xi(t), \mathbf{1}_L \rangle$$

$$\dot{\xi}(t) = \beta\text{strain}(t) - \alpha\xi(t)$$

$$\xi(0) = 0$$

6 piecewise-constant

- Exactly matches RNO model architecture

RNO architecture

$$\sigma(t) = F(\text{strain}(t), \partial_t\text{strain}(t), \xi(t))$$

$$\dot{\xi}(t) = G(\text{strain}(t), \xi(t))$$

$$\xi(0) = 0$$

Bhattacharya, K., Liu, B., Stuart, A., & **Trautner, M.** (2023). Learning Markovian homogenized models in viscoelasticity. *Multiscale Modeling & Simulation*, *21*(2), 641-679.

Caltech

RNO approximation theorem

RNO Approximation Theorem (Informal)

Let Ψ_0^{PC} be the Markovian constitutive law for 1d piecewise-constant viscoelastic materials.

For any $\eta > 0$, there exists an RNO Ψ_0^{RNO} such that

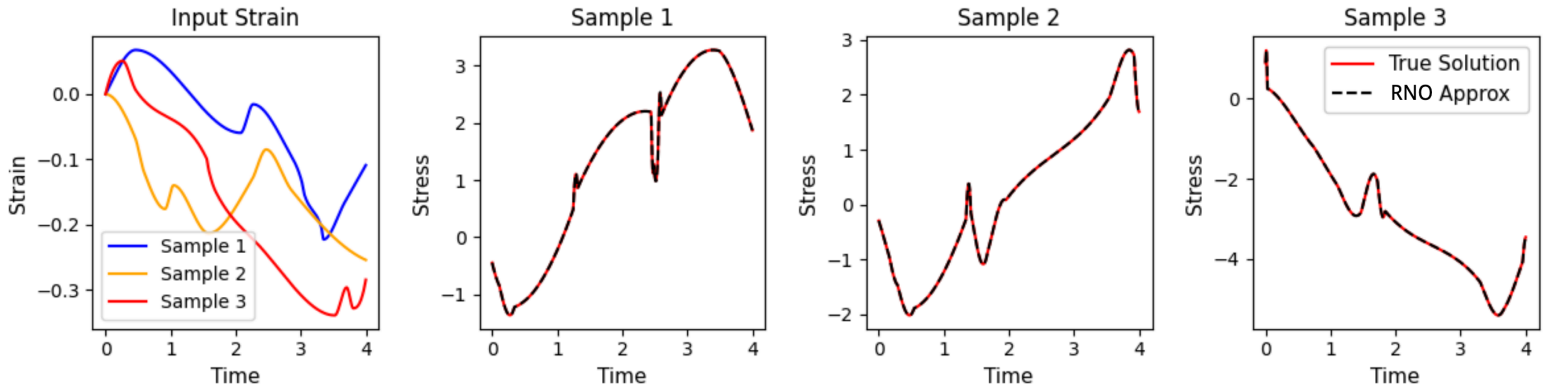
$$\sup_{t\in[0,T]} \left| \Psi_0^{\mathrm{PC}}(\{\mathrm{strain}(\tau)\}_{\tau\in[0,T]}, t) - \Psi_0^{\mathrm{RNO}}(\{\mathrm{strain}(\tau)\}_{\tau\in[0,T]}, t) \right| < \eta,$$

- Also show that u_ε for piecewise-continuous materials is approximated by u_0^{PC} for some choice of piecewise-constant materials

→ RNO is proven to approximate the history-dependent constitutive law

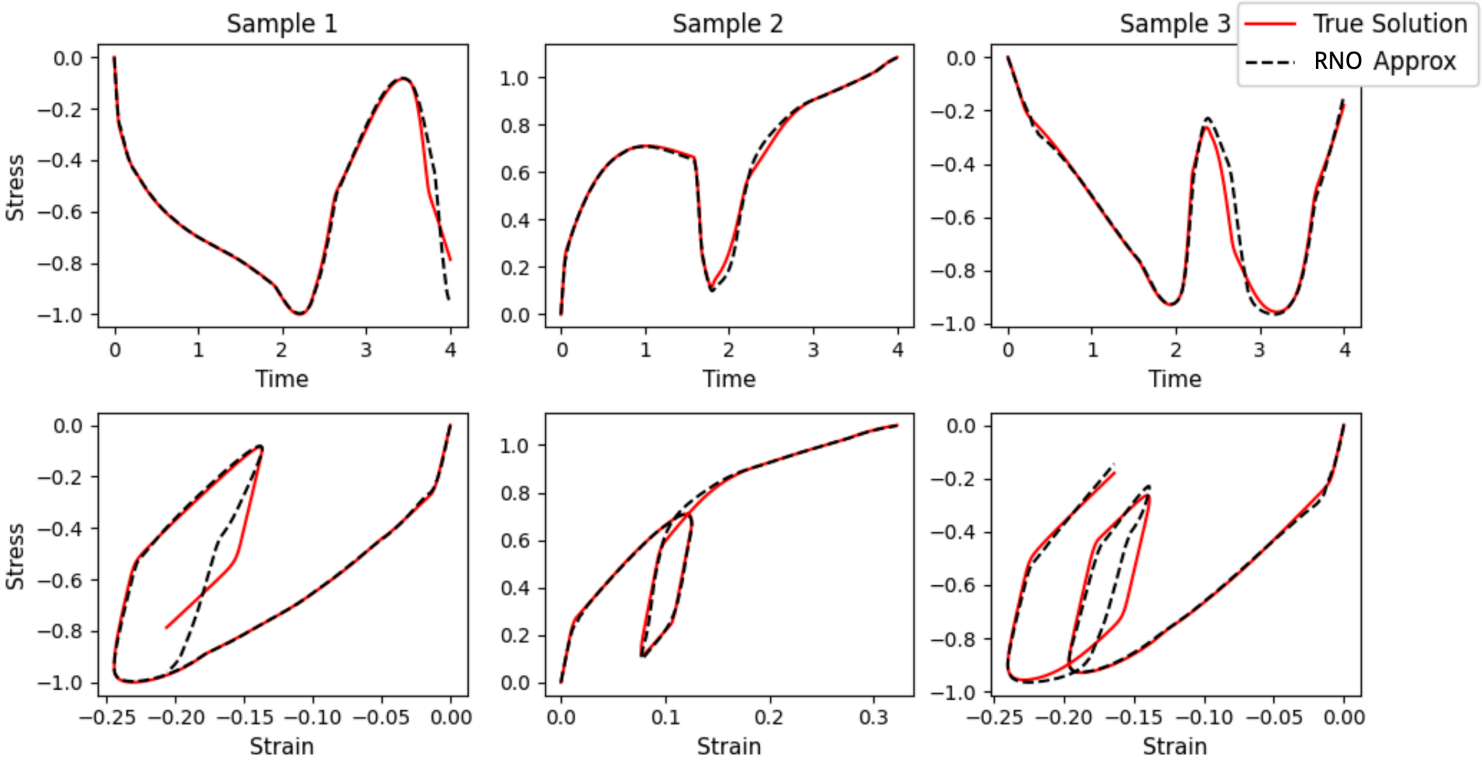
Bhattacharya, K., Liu, B., Stuart, A., & **Trautner, M**. (2023). Learning Markovian homogenized models in viscoelasticity. *Multiscale Modeling & Simulation*, *21*(2), 641-679.

Experimental results: 1d piecewise-constant viscoelasticity



Caltech

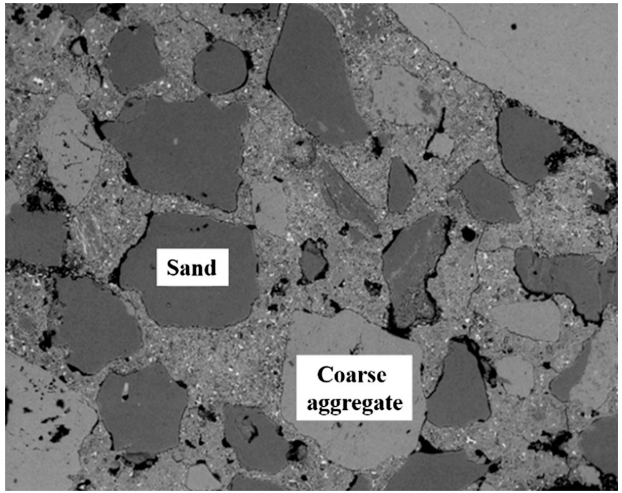
Experimental results: 1d elasto-viscoplasticity



Burigede Liu, Eric Ocegueda, **M. Trautner,** Andrew M. Stuart, and Kaushik Bhattacharya. "Learning macroscopic internal variables and history dependence from microscopic models". *Journal of the Mechanics and Physics of Solids* 178 (2023), p. 105329. doi: 10.1016/j.jmps.2023.105329.

Caltech

Material models in 2d



Grains in a cement-based material

Voronoi tessellation

Learn 2d elasticity constitutive model using a Fourier Neural Operator

Caltech

https://ascelibrary.org/doi/10.1061/%28ASCE%29MT.1943-5533.0003494

Fourier Neural Operator (FNO)

FNO Ψ is an integral kernel based neural operator architecture mapping between Banach spaces $\mathcal{A} \to \mathcal{U}$

FNO Model

$$\Psi = \mathcal{Q} \circ \mathsf{L}_{T-1} \circ \ldots \circ \mathsf{L}_0 \circ \mathcal{P}$$

Hidden FNO Layers

$$\mathsf{L}_t v_t = \sigma_t(W_t v_t + \mathcal{K}_t v_t + b_t)$$

Integral Kernel Operator

$$(\mathcal{K}_t v_t)(x) = \sum_{k \in \mathbb{Z}^d} (P_t^{(k)}) \hat{v}_t(k) e^{2\pi i \langle k, x \rangle}$$

Fourier coefficients of v_t

Caltech

Linear multiscale elasticity

$$-\nabla_x \cdot (A_\varepsilon \nabla_x u_\varepsilon) = f \qquad A_\varepsilon(x) = A\left(\frac{x}{\varepsilon}\right) \qquad A(\cdot) \quad 1-\text{periodic}$$

Homogenized equations:

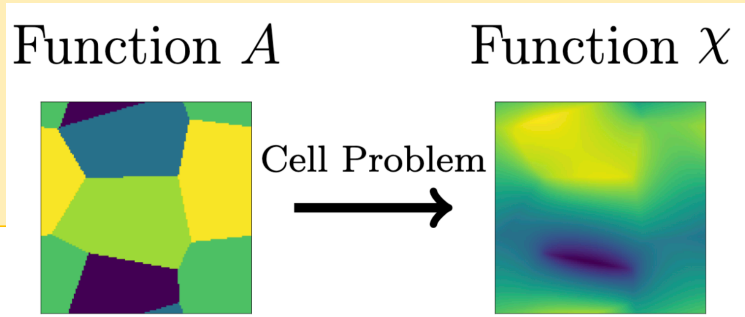
$$-\nabla_x \cdot (\overline{A}\, \nabla_x u) = f$$

$$\overline{A} = \int_{\mathbb{T}^d} (A(y) + A(y)\nabla\chi(y)^T)\, \mathrm{d}y$$

Effective coefficient

χ solves the *cell problem*

$$-\nabla \cdot (\nabla\chi A) = \nabla \cdot A \qquad \chi \quad 1-\text{periodic}$$



Interested in map $A \mapsto \chi$.

Stability and universal approximation

Goal: approximate $A \mapsto \chi$ using FNO.

Universal approximation requires
(1) a separable input space and
(2) a continuous true map.

$A \mapsto \chi$	Separable Input Space?	Continuous map?
$L^\infty \mapsto \dot{H}^1$		
$L^p \mapsto \dot{H}^1, p \in [2,\infty)$		Our result

Caltech

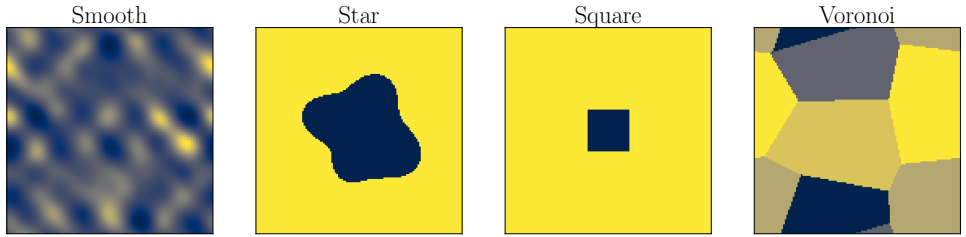
Continuity result

Cell Problem Continuity (Informal)

- $A \mapsto \chi$ is *continuous* from closed set in L^2 to $\dot{H}^1$
- There exists $q_0 > 2$ such that for all $q \in (q_0, \infty)$, $A \mapsto \chi$ is *Lipschitz continuous* from L^q to $\dot{H}^1$

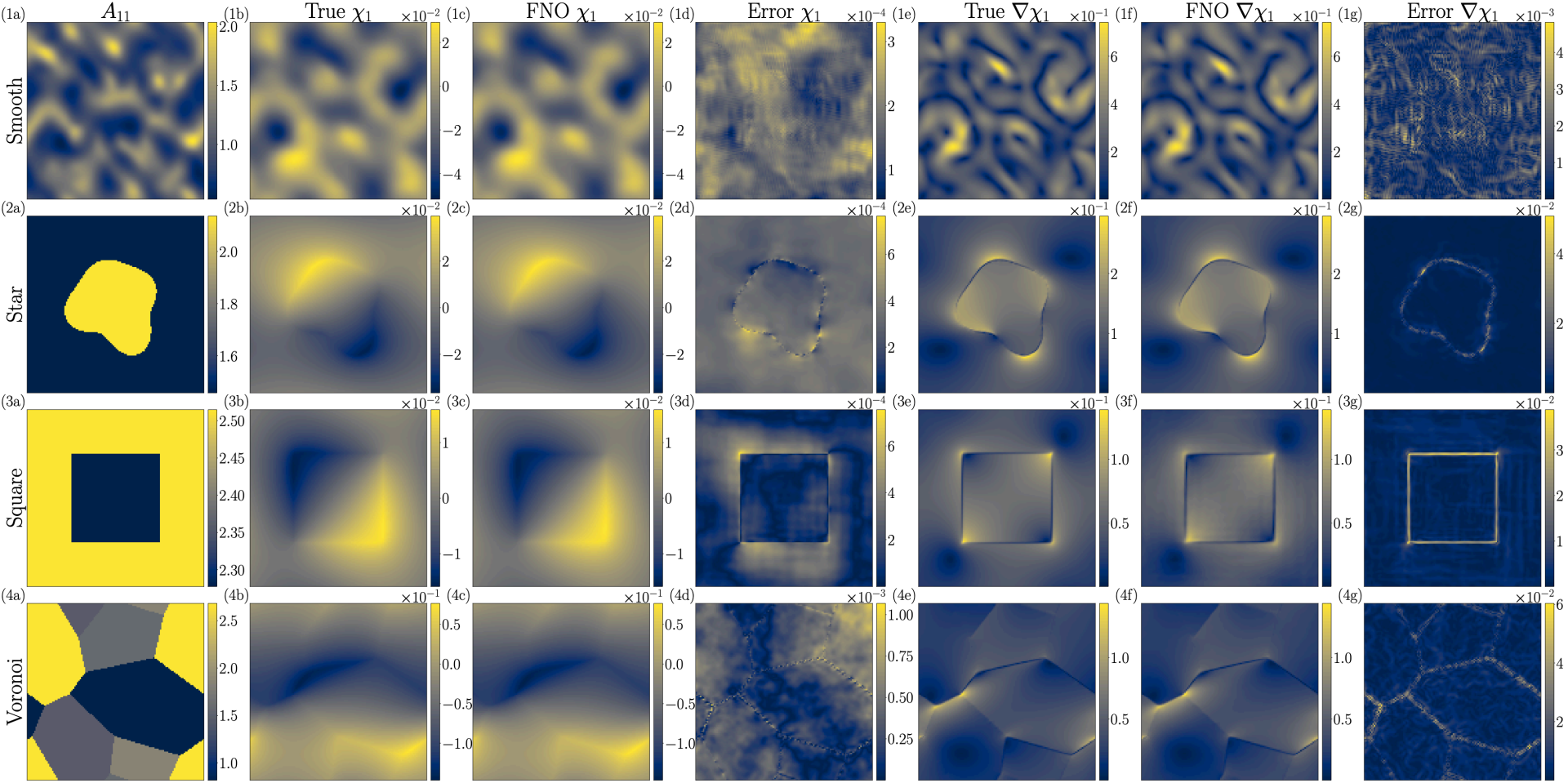
Continuity result + compactness of input set = Universal Approximation

Input set $\subset \mathrm{BV} \cap L^\infty$ is sufficient for compactness in L^2.

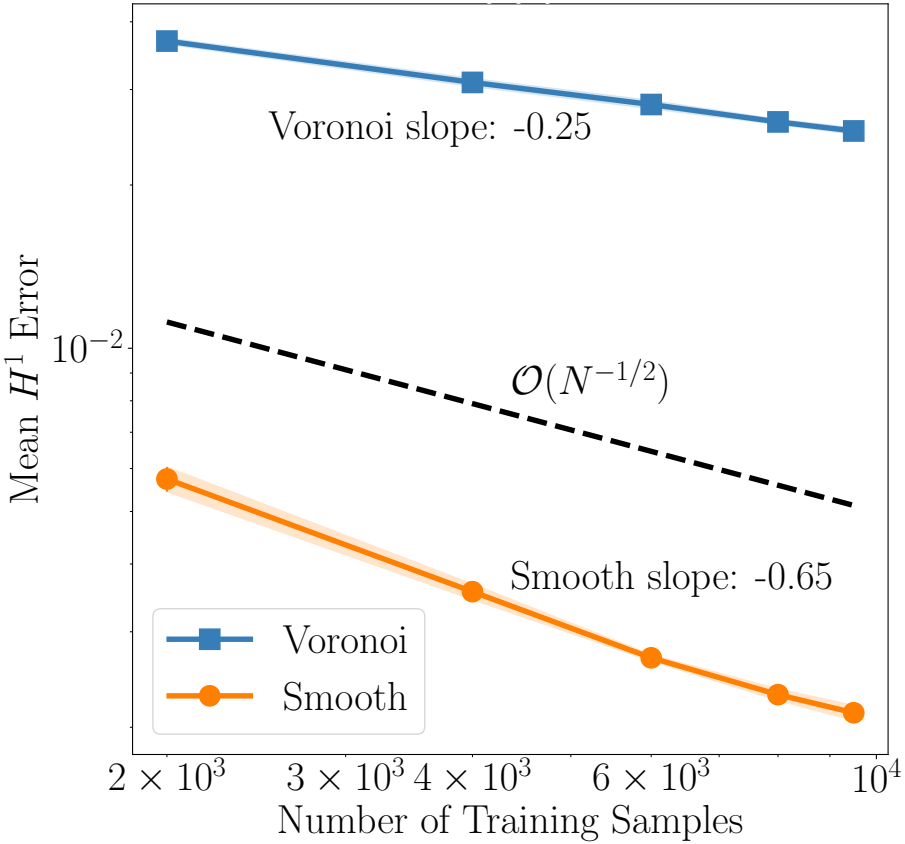


Operator learning is mathematically justified for a variety of common microstructures

Bhattacharya, K., Kovachki, N. B., Rajan, A., Stuart, A. M., & **Trautner, M.** (2024). Learning homogenization for elliptic operators. *SIAM Journal on Numerical Analysis*, *62*(4), 1844-1873.


(1a)
A_{11}
(1b)
True χ_1
$\times 10^{-2}$
(1c)
FNO χ_1
$\times 10^{-2}$
(1d)
Error χ_1
$\times 10^{-4}$
(1e)
True $\nabla\chi_1$
$\times 10^{-1}$
(1f)
FNO $\nabla\chi_1$
$\times 10^{-1}$
(1g)
Error $\nabla\chi_1$
$\times 10^{-3}$
Smooth
(2a)
(2b)
$\times 10^{-2}$
(2c)
$\times 10^{-2}$
(2d)
$\times 10^{-4}$
(2e)
$\times 10^{-1}$
(2f)
$\times 10^{-1}$
(2g)
$\times 10^{-2}$
Star
(3a)
(3b)
$\times 10^{-2}$
(3c)
$\times 10^{-2}$
(3d)
$\times 10^{-4}$
(3e)
$\times 10^{-1}$
(3f)
$\times 10^{-1}$
(3g)
$\times 10^{-2}$
Square
(4a)
(4b)
$\times 10^{-1}$
(4c)
$\times 10^{-1}$
(4d)
$\times 10^{-3}$
(4e)
(4f)
(4g)
$\times 10^{-2}$
Voronoi

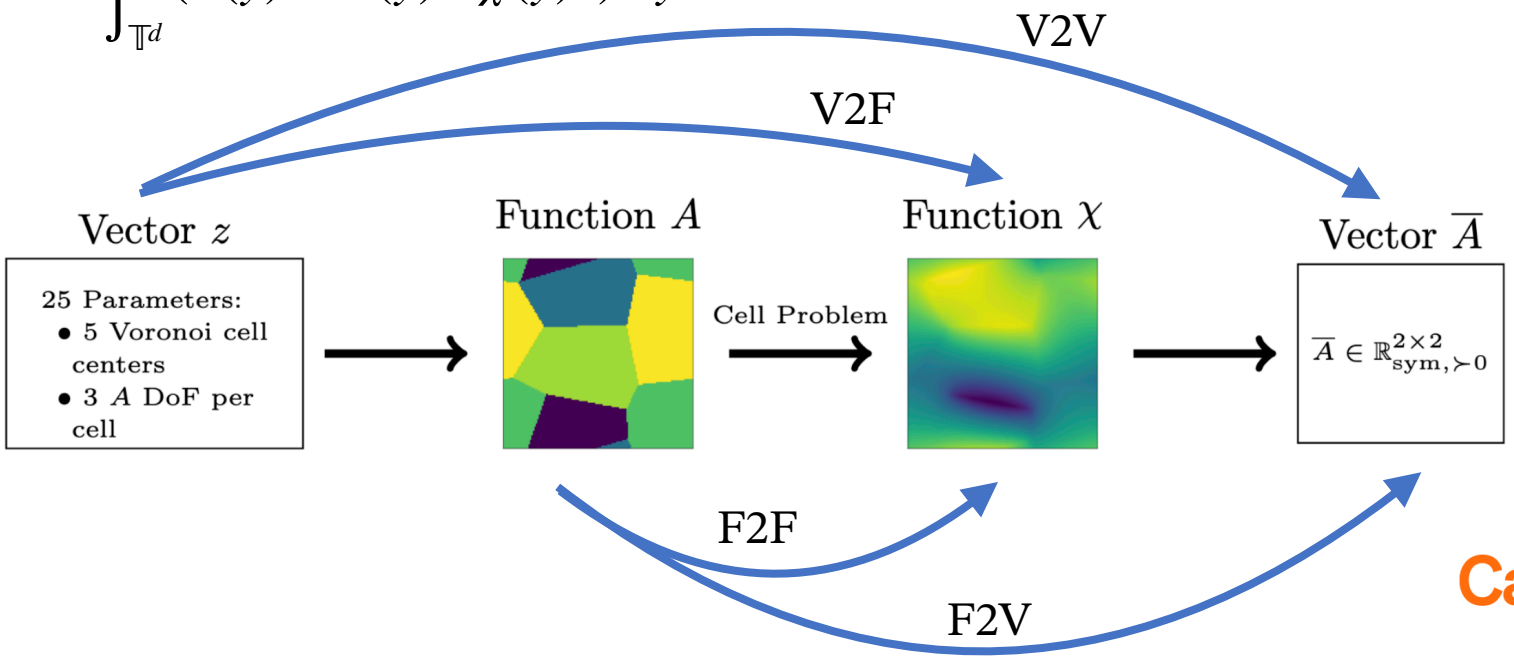
Data efficiency



Caltech

Different ways to learn effective coefficient $\overline{A}$

- Learned $A \mapsto \chi$, a map between function spaces
- Would it be more data efficient to learn effective coefficient directly, i.e. $A \mapsto \overline{A}$?
- $\overline{A} = \int_{\mathbb{T}^d} (A(y) + A(y)\nabla\chi(y)^T)\, \mathrm{d}y$



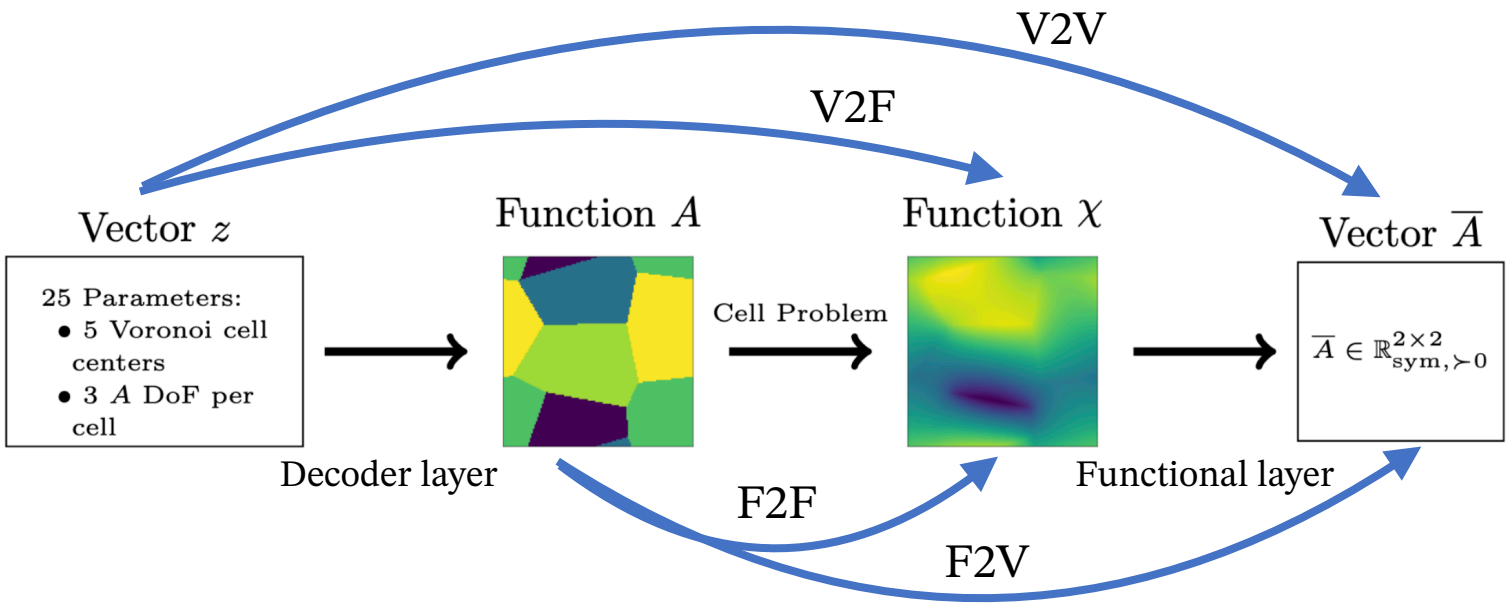
Caltech

Fourier Neural Mappings: “parameter-to-observable”

Modification of FNO to accommodate finite-dimensional inputs and outputs

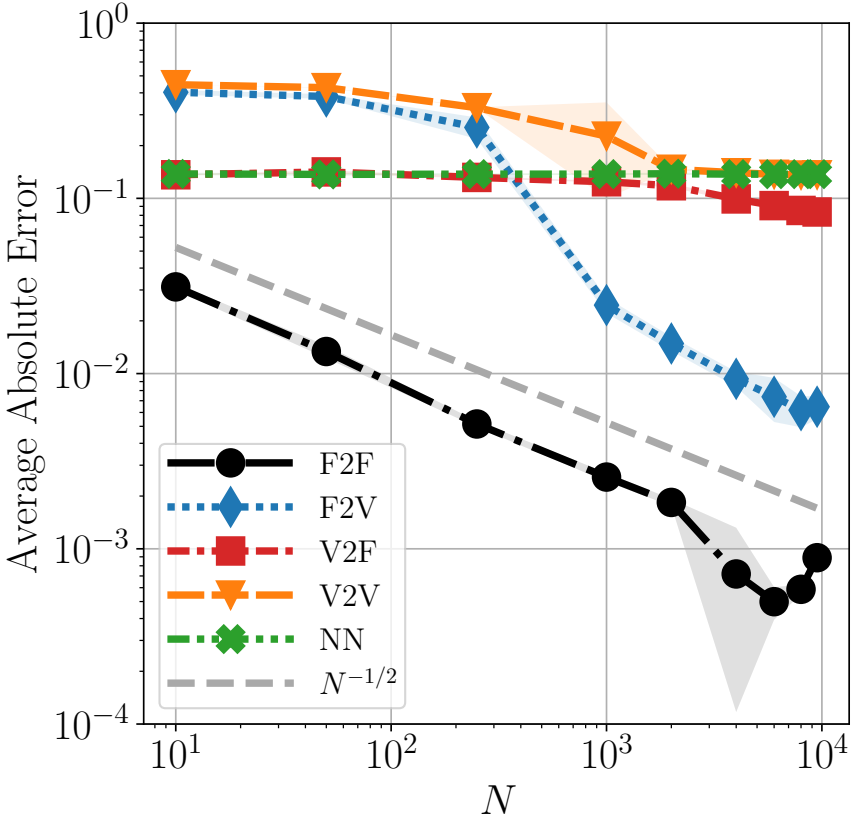
Linear *functional* layer takes functions to $\mathbb{R}^n$

Linear *decoder* layer takes $z \in \mathbb{R}^n$ to function space



Daniel Zhengyu Huang, Nicholas H. Nelsen, **M. Trautner**. An operator learning perspective on parameter-to-observable maps. Foundations of Data Science, 2025, 7(1): 163-225. doi: 10.3934/fods.2024037

Absolute error in $\|\overline{A}\|_F$ versus data size N



Valuable to use function space data if you have access to it

Learning material dependence and memory simultaneously

RNO architecture

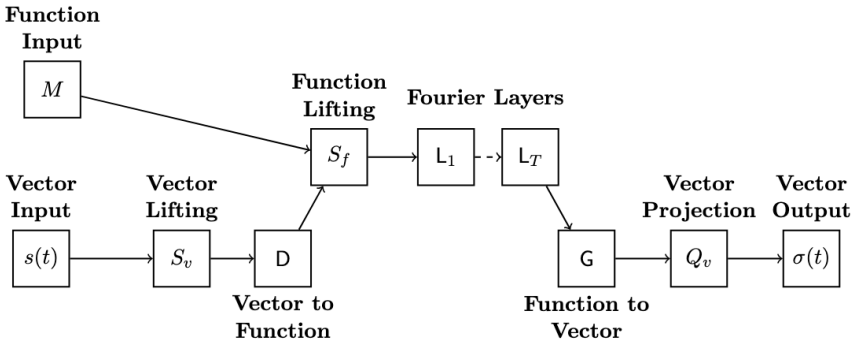
$$\sigma(t) = F(\overline{\epsilon}(t), \dot{\overline{\epsilon}}(t), \xi(t))$$
$$\dot{\xi}(t) = G(\overline{\epsilon}(t), \xi(t))$$
$$\xi(0) = 0$$

+

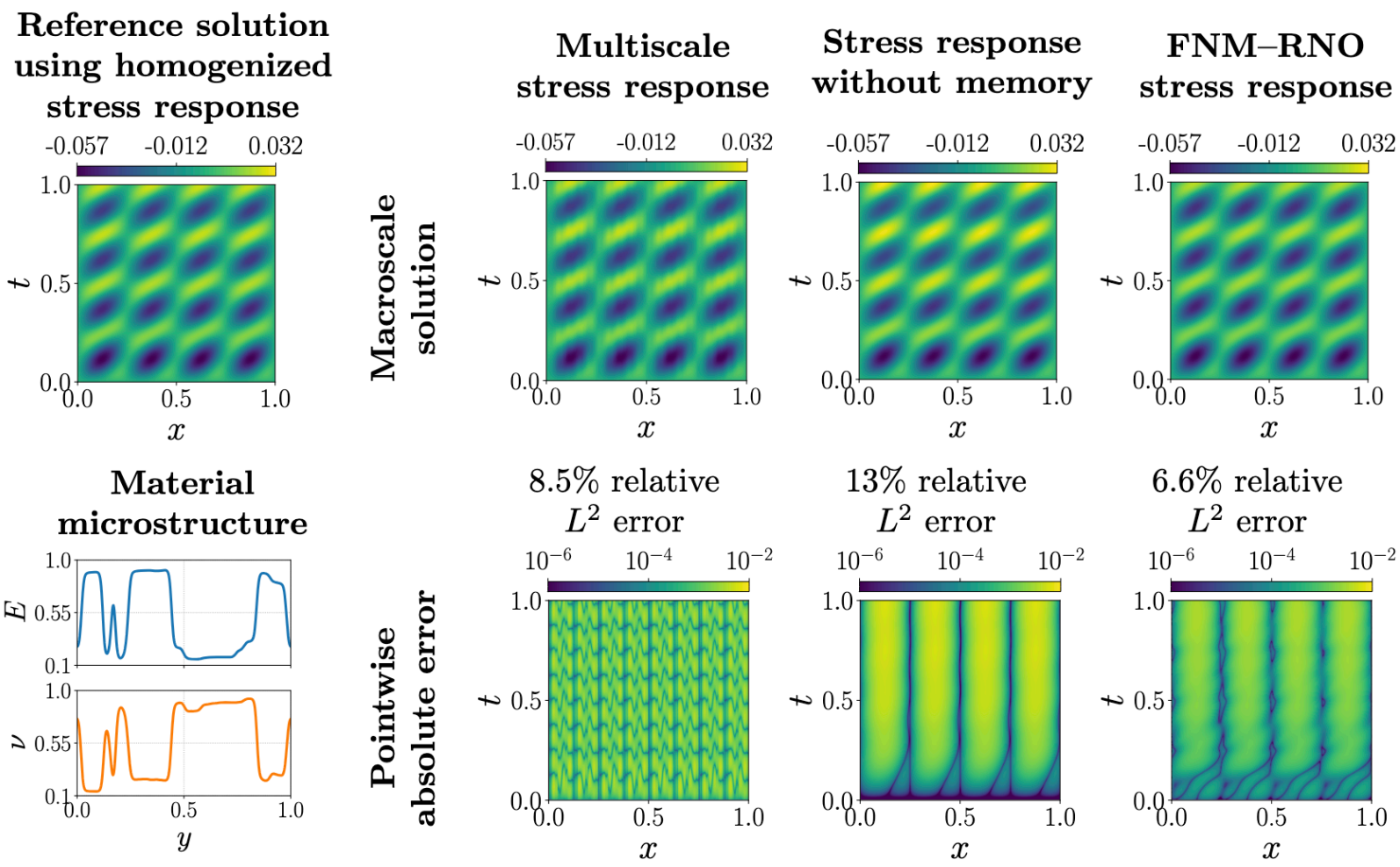
Make F and G Fourier Neural Mappings

+

→ The same model can learn history-dependence and material-dependence simultaneously.

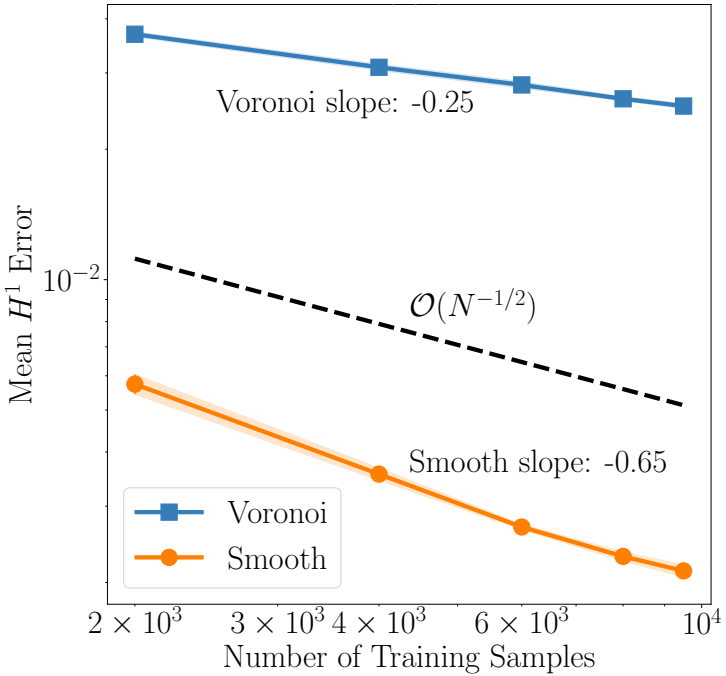


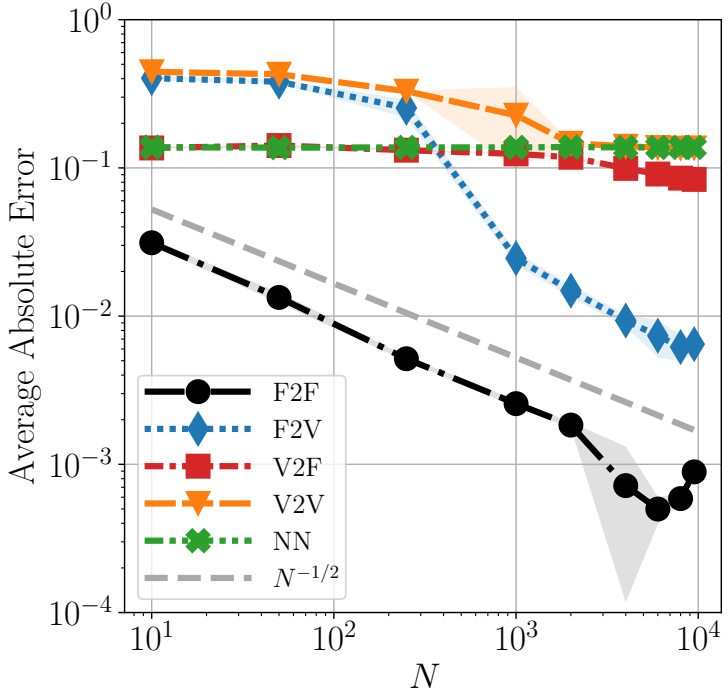
Bhattacharya, Kaushik, Lianghao Cao, George Stepaniants, Andrew Stuart, and **M. Trautner**. "Learning Memory and Material Dependent Constitutive Laws." *arXiv preprint arXiv:2502.05463* (2025).



Bhattacharya, Kaushik, Lianghao Cao, George Stepaniants, Andrew Stuart, and **M. Trautner**. "Learning Memory and Material Dependent Constitutive Laws." *arXiv preprint arXiv:2502.05463* (2025).

Sampling convergence rates- what does theory tell us?





Caltech

Theory-to-practice gap: introduction

Theory:

- Universal approximation: model expressivity
- Model size bounds: how wide and deep a model needs to be to achieve low error

Practice:

- Must identify model from finite data: data efficiency

Is there a relationship between models with high model expressivity and models with high data efficiency?

Caltech

Error convergence rates

Theory: parametric convergence rate

- A model with n parameters can approximate with error $\lesssim n^{-\alpha}$

Practice: sampling convergence rate

- A model given N data samples can find a reconstruction with error $\lesssim N^{-\beta_*}$

High α = good model expressivity

High β_* = good data efficiency

What is the relationship between α and β_*?

Caltech

Convergence rate “gap” for neural networks

High α = good model expressivity

High β_* = good data efficiency

Degrees of freedom argument suggests $\beta_* = \alpha$

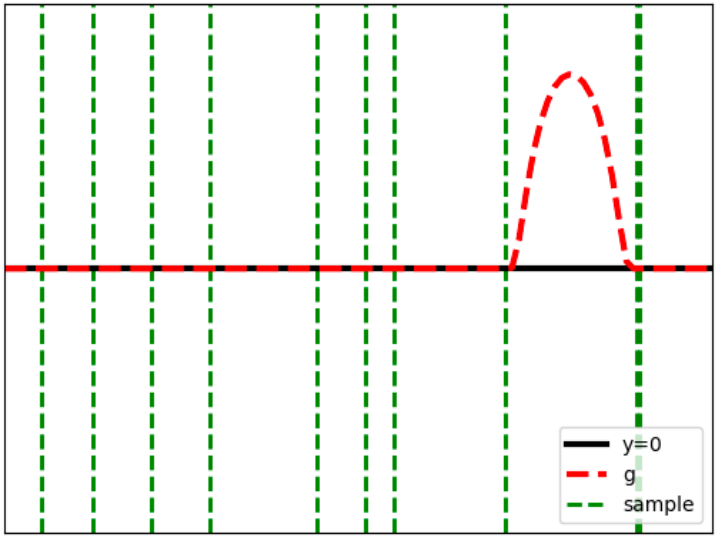
For polynomial reconstruction, determining n parameters requires $N = n + 1$ samples.

For neural networks, there is a *gap* between α and β_*.

Grohs, Philipp, and Felix Voigtlaender. "Proof of the theory-to-practice gap in deep learning via sampling complexity bounds for neural network approximation spaces." *Foundations of Computational Mathematics* 24, no. 4 (2024): 1085-1143.

Theory-to-practice gap in finite dimensions: intuition

- For any sample points $x_1, \ldots, x_N$, there exists a sufficiently large "void"
- A locally supported function g in this void is indistinguishable from the zero function
- ReLU neural networks can efficiently approximate certain locally supported functions



$$\text{good parametric complexity} \not\Longrightarrow \text{good sampling complexity}$$

Grohs, Philipp, Samuel Lanthaler, and **M. Trautner**. "Theory-to-practice gap for neural networks and neural operators." *arXiv preprint arXiv:2503.18219* (2025).

Caltech

Theory-to-practice gap in finite dimensions

α is the *parametric convergence rate*: error in the number of nonzero weights n converges at a rate $n^{-\alpha}$. High α = good model expressivity.

β_* is the *sampling convergence rate* for approximation in L^p: expected L^p error of the reconstruction with N data samples converges at a rate $N^{-\beta_*}$. High β = good data efficiency.

Theorem (Theory-to-practice gap in finite dimensions) (Informal)

For approximation of functions $[0,1]^d \mapsto \mathbb{R}$ with parametric convergence rate α,

$$\beta_* \leq \frac{1}{p} + \frac{1}{d} \cdot \frac{\alpha}{\alpha + \text{correction}^*}.$$

Grohs, Philipp, Samuel Lanthaler, and **M. Trautner**. "Theory-to-practice gap for neural networks and neural operators." *arXiv preprint arXiv:2503.18219* (2025).

Theory-to-practice gap in infinite dimensions

α is the *parametric convergence rate*: error in the number of nonzero weights n converges at a rate $n^{-\alpha}$. High α = good model expressivity.

β_* is the *sampling convergence rate* for approximation in L^p: expected L^p error of the reconstruction with N data samples converges at a rate $N^{-\beta_*}$. High β = good data efficiency.

Theorem (Theory-to-practice gap in infinite dimensions) (Informal)

For approximation of functions $\mathcal{X} \mapsto \mathbb{R}$ with parametric convergence rate α,

$$\beta_* \leq \frac{1}{p}.$$

Grohs, Philipp, Samuel Lanthaler, and **M. Trautner**. "Theory-to-practice gap for neural networks and neural operators." *arXiv preprint arXiv:2503.18219* (2025).

Conclusion

I. Recurrent neural operator captures history dependence
- Effective in various application settings in multiscale constitutive surrogate modeling.

II. In 2d, we prove that learning on discontinuous materials is theoretically valid
- Numerical experiments show that FNO can be empirically effective in this setting.
- Bound discretization error of FNO due to aliasing as well.

III. Modified the FNO to accept finite-dimensional inputs and outputs
- Data efficiency of learning in this setting is explored.
- This modification also allows for material dependence to be built into the model.

IV. Hardness result for sampling convergence rate via the lens of the “theory-to-practice gap”

Appendix

Summary of operator learning for constitutive models

I. Recurrent neural operator captures history dependence
- For 1d viscoelasticity, use is rigorously justified.
- Effective in setting of 1d elasto-viscoplasticity, more complex materials as well.

II. In 2d, we prove that learning on discontinuous materials is theoretically valid
- Numerical experiments show that FNO can be empirically effective in this setting.

III. Modified the FNO to accept finite-dimensional inputs and outputs
- Data efficiency of learning in this setting is explored.
- This modification also allows for material dependence to be built into the model.

Caltech

Discretization error in FNO (aliasing error)

In definition, $\widehat{v}_t(k) = \int_{\mathbb{T}^d} v_t(x) e^{-2\pi i \langle k,x \rangle} \, \mathrm{d}x$

In practice: $\widehat{v}_t^N(k) = \mathsf{DFT}(v_t^N)$

$$\Psi = \mathcal{Q} \circ \mathsf{L}_{T-1} \circ \ldots \circ \mathsf{L}_0 \circ \mathcal{P}$$

$$\mathsf{L}_t v_t = \sigma_t(W_t v_t + \mathcal{K}_t v_t + b_t)$$

$$(\mathcal{K}_t v_t)(x) = \sum_{k \in \mathbb{Z}^d} (P_t^{(k)}) \widehat{v}_t(k) e^{2\pi i \langle k,x \rangle}$$

Theorem (Informal) (Bounding FNO discretization error)

For input a with Sobolev regularity s, the error evaluated on the grid $[N]^d$ of the output of layer L_t satisfies

$$\frac{1}{N^{d/2}} \| v_t(a) - v_t^N(a) \|_{\ell^2([N]^d)} \leq C N^{-s}$$

Lanthaler, Samuel, Andrew M. Stuart, and **M. Trautner**. "Discretization error of Fourier neural operators." *arXiv preprint arXiv:2405.02221* (2024).

Setting: multiscale material response

Displacement — Stress — External forcing

Force balance $$\rho \partial_t^2 u_\varepsilon(x,t) = \nabla_x \cdot \sigma_\varepsilon(x,t) + f(x,t),$$

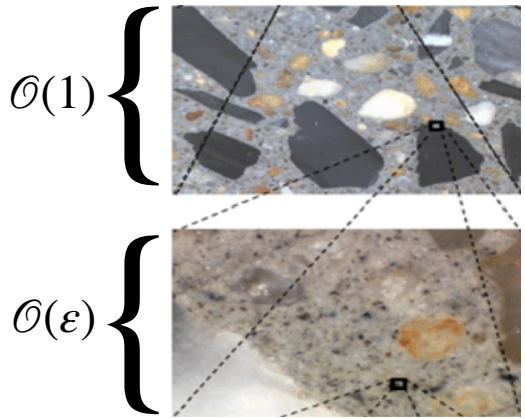
Constitutive law $$\sigma_\varepsilon(x,t) = \Psi_\varepsilon^\dagger\big(\{\nabla_x u_\varepsilon(x,s)\}_{s\in[0,t]}; M, x\big)(t),$$

Initial condition $$u_\varepsilon(x,0) = \partial_t u_\varepsilon(x,0) = 0,$$

Boundary condition $$u_\varepsilon(x,t) = 0, \quad x \in \partial\mathcal{D}.$$

Strain history

ε indicates dependence on small scale variable $y = \dfrac{x}{\varepsilon}$



Zhang, Mingzhong, 2013.

Homogenized equations

Displacement Stress External forcing

Force balance $\rho \partial_t^2 u_0(x,t) = \nabla_x \cdot \sigma_0(x,t) + f(x,t),$

Constitutive law $\sigma_0(x,t) = \Psi_0^\dagger\big(\{\nabla_x u_0(x,s)\}_{s\in[0,t]}; M\big)(t),$

Initial condition $u_0(x,0) = \partial_t u_0(x,0) = 0,$

Strain history

Boundary condition $u_0(x,t) = 0, \quad x \in \partial\mathcal{D}.$

- Assume microstructure ε periodic
- Seek *homogenized constitutive law* $\Psi_0^\dagger$ such that $u_\varepsilon \to u_0$ as $\varepsilon \to 0$
- $\Psi_0^\dagger$: homogenized strain $\mapsto$ homogenized stress

Caltech

History-dependent constitutive laws

- For some materials, stress depend on entire strain history

Kelvin-Voigt viscoelasticity

Strain: $\epsilon_\varepsilon = \nabla_x u_\varepsilon$ Homogenized strain: $\bar{\epsilon} = \int_\Omega \nabla_x u_\varepsilon \, \mathrm{d}x$

Multiscale constitutive law $\sigma_\epsilon(t) = E_\varepsilon \epsilon_\varepsilon(t) + \nu_\varepsilon \dot{\epsilon}_\varepsilon(t)$

Homogenized constitutive law $\sigma(t) = E'\bar{\epsilon}(t) + \nu'\dot{\bar{\epsilon}}(t) - \int_0^t \kappa(t-\tau)\bar{\epsilon}(\tau) \, \mathrm{d}\tau$

$\Psi_0^\dagger : \{\bar{\epsilon}(\tau)\}_{\tau\in[0,T]} \mapsto \{\sigma(\tau)\}_{\tau\in[0,T]}$

History dependence

Caltech

RNO approximation theorem

RNO Approximation Theorem (Informal)

Let Ψ_0^{PC} be the Markovian form taken by $\Psi_0^{\dagger}$ for 1d PC viscoelastic materials.

For any $\eta > 0$, there exists Ψ_0^{RNO} such that

$$\sup_{t\in[0,T];\epsilon\in\mathsf{Z}} \left| \Psi_0^{\mathsf{PC}}(\{\epsilon(\tau)\}_{\tau\in[0,T]}, t) - \Psi_0^{\mathsf{RNO}}(\{\epsilon(\tau)\}_{\tau\in[0,T]}, t) \right| < \eta,$$

where Z bounds ϵ uniformly in t.

- Also show that u_ε for piecewise-continuous materials is approximated by u_0^{PC} for some choice of PC materials
- Francfort & Suquet (1986) gives $u_\varepsilon \to u_0$

Bhattacharya, K., Liu, B., Stuart, A., & **Trautner, M.** (2023). Learning Markovian homogenized models in viscoelasticity. *Multiscale Modeling & Simulation*, *21*(2), 641-679.

Homogenization two ways

1. We have defined homogenized map $\Psi_0^\dagger : \{\overline{\epsilon}(\tau)\}_{0\le\tau\le t} \mapsto \{\sigma(\tau)\}_{0\le\tau\le t}$ where $\overline{\epsilon} = \nabla_x u_0$ as the map from the homogenized strain trajectory to the homogenized stress trajectory
2. Another perspective views $\Psi_0^\dagger$ as the map from spatially-averaged strain trajectory over the cell to spatially averaged stress trajectory over the cell, i.e. $\int_{\mathbb{T}^d} \nabla_x u_\varepsilon(y)\, \mathrm{d}y$ to $\int_{\mathbb{T}^d} \sigma_\varepsilon(y)\, \mathrm{d}y$.

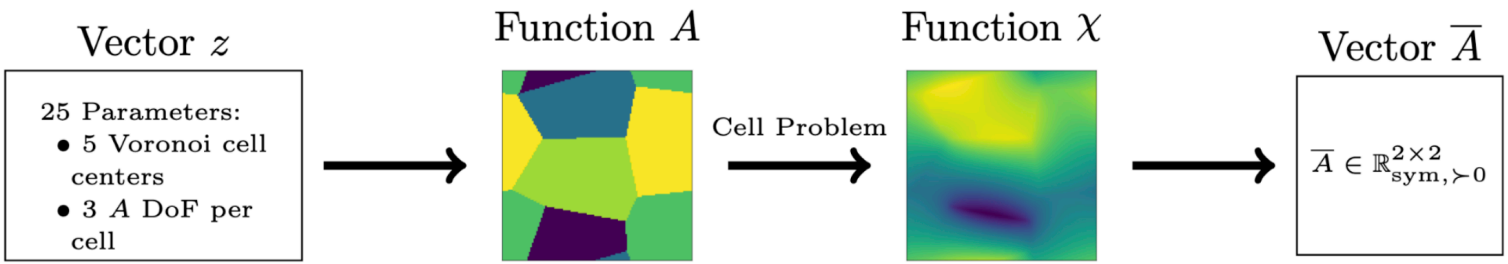
Caltech

Fourier Neural Mappings

- Modification of FNO to accommodate finite-dimensional inputs and outputs with underlying infinite-dimensional operator map

Linear *functional* layer $h \mapsto \mathcal{G}h := \int_\Omega \kappa(x)h(x)\,\mathrm{d}x$ takes functions to $\mathbb{R}^n$

Linear *decoder* layer $z \mapsto \mathcal{D}z := \kappa(\,\cdot\,)z$ takes $z \in \mathbb{R}^n$ to function space



Daniel Zhengyu Huang, Nicholas H. Nelsen, **Margaret Trautner**. An operator learning perspective on parameter-to-observable maps. Foundations of Data Science, 2025, 7(1): 163-225. doi: 10.3934/fods.2024037

Theory-to-practice gap: introduction

Many theoretical results in machine learning answer questions about *model expressivity*:

- Can this model class approximate to ε accuracy any function in a certain set? (*Universal approximation*)
- Given n nonzero model weights, how small is the error of the optimal model parameterization? (*Parametric complexity*)

Other results answer questions about *generalizability*:

- Given N data samples from the underlying map of interest, how well does Algorithm X do, on average, compared to the optimal model parameterization? (*Optimization*)
- Given N data samples, how well can *any* algorithm reconstruct the true underlying map of interest? (*Sampling complexity*)

Caltech

Infinite-dimensional setting

	Finite-Dimensional Setting	Infinite-Dimensional Setting
Mapping between	$\mathbb{R}^d \to \mathbb{R}$	$\mathcal{X} \to \mathbb{R}$ for separable Banach space $\mathcal{X}$
Restriction of input space	$D = [0,1]^d \subset \mathbb{R}^d$	Measure on $\mathcal{X}$
Expected error measured in	$L^p(D)$	Banach space V such that $U \subset V$

More formal set-up: Neural Operator Approximation Spaces

- Approximation Space $A^\alpha = \{\mathcal{G} \in C(\mathcal{X}; \mathcal{Y}) : \|\mathcal{G}\|_{A^\alpha} < \infty\}$
- Approximation space quasi-norm $\|\mathcal{G}\|_{A^\alpha} = \inf\{\theta > 0 : \Gamma^\alpha(\mathcal{G}/\theta) \leq 1\}$
- $\Gamma^\alpha(\mathcal{G}) = \max\{\sup_{u\in\mathcal{X}} \|\mathcal{G}(u)\|, \sup_{n\in\mathbb{N}}[n^\alpha \cdot \inf_{\Psi\in\Sigma_n} \sup_{u\in\mathcal{X}} \|\mathcal{G}(u) - \Psi(u)\|_{\mathcal{Y}}]\}$
- Σ_n contains neural operators with at most n nonzero weights of uniformly bounded magnitude

A^α contains all functions $\mathcal{X} \to \mathcal{Y}$ that can be uniformly approximated at rate $n^{-\alpha}$ by neural operators with at most n nonzero weights (parameters) of bounded magnitude

- We work in unit ball $U^\alpha \subset A^\alpha$

Caltech

More formal set-up: Algorithms

- $\mathrm{Alg}_N(U, V)$ set of all deterministic methods using N point measurements
- $\mathscr{A} \in \mathrm{Alg}_N(U, V)$ if there exist samples $(x_1, \ldots, x_N) \in \mathscr{X}^N$ and map $Q : \mathbb{R}^N \to V$ such that $\mathscr{A}(G) = Q(G(x_1), \ldots, G(x_N))$ for all $G \in U$
- Optimal error $e_N = \inf_{\mathscr{A} \in \mathrm{Alg}_N} \sup_{G \in U} \|\mathscr{A}(G) - G\|_V$
- Can also formulate set of randomized algorithms
- Results hold over all reconstruction methods using N samples

Caltech

Proof sketch

1. Show that approximation space $U \subset C(\mathcal{X}, \mathbb{R})$ contains compositions of the form $f \circ \mathscr{E}$, where $\mathscr{E} : \mathcal{X} \to \mathbb{R}^d$ is an encoder, and $f : \mathbb{R}^d \to \mathbb{R}$ belongs to the d-dimensional approximation space $f \in U_\ell^{\alpha,\infty}$.
2. Then $f \mapsto f \circ \mathscr{E}$ defines an embedding of $U_\ell^{\alpha,\infty}$ into U, so any d-dimensional bound on the convergence rate applies in the infinite-dimensional setting.

Caltech

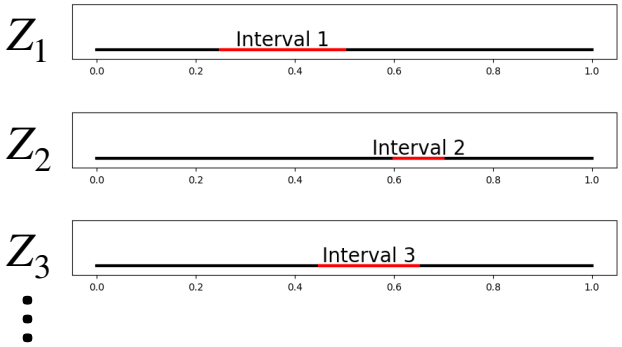
Grohs, Philipp, Samuel Lanthaler, and **Margaret Trautner**. "Theory-to-practice gap for neural networks and neural operators." *arXiv preprint arXiv:2503.18219* (2025).

Proof sketch

Require measure μ on $\mathcal{X}$ and encoder $\mathcal{E} : \mathcal{X} \mapsto \mathbb{R}^d$ such that $\mathcal{E}$ "fills out" the unit cube $[0,1]^d$ in the following sense:

$$\mathcal{E}_{\#}\mu \geq c \cdot \mathrm{Unif}([0,1]^d)$$

Assume $\mu \in \mathcal{P}(\mathcal{X})$ can be written as the law of $u = \sum_{j=1}^{\infty} Z_j e_j$ for biorthogonal sequences $\{e_j\}_{j\in\mathbb{N}} \subset \mathcal{X}$, $\{e_j^*\}_{j\in\mathbb{N}} \subset \mathcal{X}^*$ and iid real-valued random variables Z_j with sufficient density on open intervals.



By approximating a projection onto $\{e_j\}_{j\in[d]}$, the encoder "fills out" a cube with sufficient density that can be rescaled to $[0,1]^d$

Caltech

Grohs, Philipp, Samuel Lanthaler, and **Margaret Trautner**. "Theory-to-practice gap for neural networks and neural operators." *arXiv preprint arXiv:2503.18219* (2025).