Learning Solution Operators For PDEs

Algorithms, Analysis and Applications

Andrew Stuart

Computing and Mathematical Sciences
California Institute of Technology

Various Locations in Paris

December 14th and 15th, 2023
Table of Contents

Set-Up

Algorithms

Analysis

Application I: Material-Dependent Constitutive Models

Application II: History-Dependent Constitutive Models

Conclusions
Talk Outline

Set-Up

Algorithms

Analysis

Application I: Material-Dependent Constitutive Models

Application II: History-Dependent Constitutive Models

Conclusions
Operator Learning

Supervised Learning

Determine \( \Psi^\dag : \mathcal{U} \to \mathcal{V} \) from samples

\[
\{ u_n, \Psi^\dag (u_n) \}_{n=1}^N, \quad u_n \sim \mu.
\]

Probability measure \( \mu \) supported on \( \mathcal{U} \).

In standard supervised learning \( \mathcal{U} = \mathbb{R}^{dx} \) and \( \mathcal{V} = \mathbb{R}^{dy} \) (regression) or \( \mathcal{V} = \{1, \cdots K\} \) (classification).

Supervised Learning Of Operators

Separable Banach spaces \( \mathcal{U}, \mathcal{V} \) of vector-valued functions:

\[
\mathcal{U} = \{ u : D_u \to \mathbb{R}^{d_u} \}, \quad D_u \subseteq \mathbb{R}^{d_u}
\]

\[
\mathcal{V} = \{ v : D_v \to \mathbb{R}^{d_v} \}, \quad D_v \subseteq \mathbb{R}^{d_v}.
\]
Consider a family of parameterized functions from $\mathcal{U}$ into $\mathcal{V}$:

$$\Psi : \mathcal{U} \times \Theta \mapsto \mathcal{V}.$$ 

Here $\Theta \subseteq \mathbb{R}^p$ denotes the parameter space.

$$\mu_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{u_n}, \quad \text{RE}(v, w) = \frac{\|v - w\|_\mathcal{V}}{\max\{1, \|v\|_\mathcal{V}\}},$$

$$\theta^* = \arg\min_{\theta \in \Theta} \mathcal{R}_N(\theta), \quad \mathcal{R}_N(\theta) := \mathbb{E}_{u \sim \mu_N} \text{RE}(\Psi^\dagger(u), \Psi(u; \theta)).$$
### Example (Porous Medium Flow)

**Darcy Law**

<table>
<thead>
<tr>
<th>Mass conservation</th>
<th>$-\nabla \cdot (a \nabla v) = f$, $x \in D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundary condition</td>
<td>$v = 0$, $x \in \partial D$</td>
</tr>
</tbody>
</table>

**Operator Of Interest**

| Parametric Dependence | $\psi^\dagger : a \mapsto v$ |
Example (Porous Medium Flow)

Input-Output

Input: \( a \in L^\infty(D) \) \textbf{(Left)},

Output: \( v \in H^1_0(D) \). \textbf{(Right)},
Example (Porous Medium Flow): **Discretize Then Learn**

Zhu and Zabarás 2018 [21], Khoo et al [8]

Example (Porous Medium Flow): **Learn Then Discretize**

Bhattacharya et al 2021 [2]
Examples: (Homogenized Constitutive Models)

Material Properties Dependence

- $\mathcal{U}$: material properties $A(\cdot)$.
- $\mathcal{V}$: stress $\sigma$.
- $\sigma = \Psi^\dagger(A)(\nabla u + \nabla u^\top)$.
- Approximate $\Psi^\dagger \approx \Psi(\cdot; \theta^*)$

History Dependence

- $\mathcal{U}$: time histories of strain $\{\nabla u\}$.
- $\mathcal{V}$: time histories of stress $\{\sigma\}$.
- $\{\sigma\} = \Psi^\dagger(\{\nabla u\})$.
- Approximate $\Psi^\dagger \approx \Psi(\cdot; \theta^*)$
Talk Outline

Set-Up

Algorithms

Analysis

Application I: Material-Dependent Constitutive Models

Application II: History-Dependent Constitutive Models

Conclusions
Finding Latent Structure

In A Picture

\[ G_U \circ F_U \approx \text{Id} \]
\[ G_V \circ F_V \approx \text{Id} \]
\[ G_V \circ \varphi \circ F_U \approx \Psi^\dagger \]
**Architecture**  Bhattacharya, Hosseini, Kovachki and AMS ’19 [2]

$$\Psi_{PCA}(u; \theta)(x) = \sum_{j=1}^{m} \alpha_j(Lu; \theta)\psi_j(x), \quad \forall u \in \mathcal{U} \quad x \in D_v.$$ 

**Details**

- \(\{\phi_j\}\) are PCA basis functions under \(\mu\).
- \(Lu = \{\langle \phi_j, u \rangle\}\) maps to PCA coefficients under \(\mu\).
- \(\{\psi_j\}\) are PCA basis functions under \((\Psi^\dagger)^\#\mu\).
- \(\{\alpha_j\}\) are finite dimensional neural networks.
**DEEPPONET**  Shallow approximation of functionals (Chen and Chen [5]) made deep

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Lu, Jin, Pang, Zhang, and Karniadakis’19 [18]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Psi_{DEEP}(u; \theta)(x) = \sum_{j=1}^{m} \alpha_j(Lu; \theta_\alpha)\psi_j(y; \theta_\psi), \quad \forall u \in \mathcal{U} \quad x \in D_\nu. )</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>▶ <em>Lu</em> maps to linear functionals on ( \mathcal{U} ).</td>
</tr>
<tr>
<td>▶ e.g. PCA coefficients under ( \mu ); or pointwise ( {u(x_\ell)} ).</td>
</tr>
<tr>
<td>▶ ( {\alpha_j, \psi_j} ) are finite dimensional neural networks.</td>
</tr>
<tr>
<td>▶ ( \theta = (\theta_\alpha, \theta_\psi) ).</td>
</tr>
</tbody>
</table>
Fourier Neural Operator (FNO) \[ \text{DNN (Goodfellow et al [7]) Extended to Operators} \]

### Architecture

\[ U = \mathcal{V} \text{ Hilbert} \]

\[ \Psi_{\text{FNO}}(u; \theta) = Q \circ \mathcal{L}_L \circ \cdots \mathcal{L}_2 \circ \mathcal{L}_1 \circ \mathcal{R}(u), \forall u \in U, \]

\[ \mathcal{L}_i(v)(x; \theta) = \sigma(W_i v(x) + b_i + \mathcal{K}(v)(x; \gamma_i)), \]

\[ \mathcal{K}(v)(x; \gamma) = \sum_{m=1}^{M} \alpha_m(\gamma(m)) \langle f_m, v \rangle_{\mathcal{U} \otimes d_c} g_m(x). \]

### Details

- \( Q, R \) pointwise NNs or linear transformations.
- \((W_i, b_i)\) define pointwise affine transformations.
- \( \mathcal{K} \) eg FFT as convolutional integral operator.
- \( \theta \) collects parameters from previous three bullets.
Talk Outline

Set-Up

Algorithms

Analysis

Application I: Material-Dependent Constitutive Models

Application II: History-Dependent Constitutive Models

Conclusions
Universal Approximation (Latent) over compact sets

Definition

A Banach space has the **approximation property (AP)** if every compact operator is a limit of finite-rank operators.

Theorem 1  Kovacvkhki '22 [10, 11]

Assume

- $\mathcal{U}, \mathcal{V}$ Banach spaces with the AP.
- $\Psi^\dagger: \mathcal{U} \to \mathcal{V}$ continuous, $K \subset \mathcal{U}$ compact.

For any $\epsilon > 0$ there exist bounded linear $F_U: \mathcal{U} \to \mathbb{R}^{d_U}$, $G_V: \mathbb{R}^{d_V} \to \mathcal{V}$, and a continuous map $\varphi \in C(\mathbb{R}^{d_U}; \mathbb{R}^{d_V})$ such that

$$
\sup_{u \in K} \| \Psi^\dagger(u) - (G_V \circ \varphi \circ F_U)(u) \|_{\mathcal{V}} \leq \epsilon.
$$
Theorem 2 Kovacvhi '22 [10, 11]

Assume

- $\mathcal{U}$ Banach space with AP, $\mathcal{V}$ separable Hilbert space.
- $\mu$ probability measure on $\mathcal{U}$.
- $\Psi^\dagger \in L^p_\mu(\mathcal{U}; \mathcal{V})$ for $1 \leq p < \infty$.

For any $\epsilon > 0 \exists$ bounded linear $F_\mathcal{U} : \mathcal{U} \to \mathbb{R}^{d_\mathcal{U}}$, $G_\mathcal{V} : \mathbb{R}^{d_\mathcal{V}} \to \mathcal{V}$, and a continuous map $\varphi \in C(\mathbb{R}^{d_\mathcal{U}} ; \mathbb{R}^{d_\mathcal{V}})$ such that

$$\|\Psi^\dagger - G_\mathcal{V} \circ \varphi \circ F_\mathcal{U}\|_{L^p_\mu(\mathcal{U}; \mathcal{V})} \leq \epsilon.$$
Universal Approximation (FNO)

**Theorem 3** Lanthaler, Li and AMS ’23 [13]

Assume
- \( \mathcal{U} = C(\overline{D}, \mathbb{R}^d) \).
- \( \mathcal{V} = C(\overline{D}, \mathbb{R}^{d'}) \).
- \( \Psi^\dagger: \mathcal{U} \to \mathcal{V} \) continuous, \( K \subset \mathcal{U} \) compact.

For any \( L, M > 0 \) and any \( \epsilon > 0 \) \( \exists \) an FNO \( \Psi(\cdot; \theta^*) : \mathcal{U} \to \mathcal{V} \) such that

\[
\sup_{u \in K} \| \Psi^\dagger(u) - \Psi(u; \theta) \|_{\mathcal{V}} \leq \epsilon.
\]

Kovachki, Lanthaler and Mishra ’21 [9] (Complexity, FNO)
Talk Outline

Set-Up

Algorithms

Analysis

Application I: Material-Dependent Constitutive Models

Application II: History-Dependent Constitutive Models

Conclusions
Multiscale Problem

Canonical Elliptic Multiscale Problem

\[-\nabla \cdot (A^\varepsilon \nabla u^\varepsilon) = f, \quad x \in \Omega\]
\[u^\varepsilon = 0, \quad x \in \partial \Omega\]
\[A^\varepsilon(x) = A \left( \frac{x}{\varepsilon} \right), \quad A \in \mathcal{U} := L^\infty(\mathbb{T}^d, \mathbb{R}^{d \times d}).\]

Standing Assumption on $A$

\[PD_{\alpha,\beta} = \{A \in L^\infty(\mathbb{T}^d; \mathbb{R}^{d \times d}) : \forall (y, \xi) \in \mathbb{T}^d \times \mathbb{R}^d, \alpha |\xi|^2 \leq \langle \xi, A(y)\xi \rangle \leq \beta |\xi|^2\}.\]
Operator Learning

Homogenized Elliptic Problem; \( u_0 \approx u^\varepsilon \)

\[-\nabla \cdot (A_0 \nabla u_0) = f, \quad x \in \Omega\]

\[u_0 = 0, \quad x \in \partial \Omega\]

\[A_0(x) = A_0, \quad \text{constant.}\]

Constitutive Model  
Bensoussan, Lions, Papanicolaou ’78 [1], Pavliotis and AMS ’08 [20][Ch12]

\( A_0 \) determined by \( \chi \in \mathcal{V} := H^1_{\text{per}}(\mathbb{T}^d, \mathbb{R}^d) \)

\[-\nabla_y \cdot (\nabla_y \chi A) = \nabla_y \cdot A, \quad y \in \mathbb{T}^d,\]

\[A_0 = \int_{\mathbb{T}^d} \left( A(y) + A(y) \nabla \chi(y)^T \right) \, dy.\]
Universal Approximation (Cell Problem Solution Operator)

Goal: Supervised Learning (FNO)
Bhattacharya, Kovachki, Rajan, AMS, Trautner ’23 [3]

- Learn map $A(\cdot) \in \mathcal{U} \mapsto \chi(\cdot) \in \mathcal{V} = \dot{H}^1_{\text{per}}(\mathbb{T}^d, \mathbb{R}^d)$.
- How to choose $\mathcal{U}$?

Theorem 4 Bhattacharya, Kovachki, Rajan, AMS, Trautner ’23 [3]

Define the mapping $\Psi^\dagger : PD_{\alpha,\beta} \to \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d)$ from the solution map $A \mapsto \chi$ given by

$$-\nabla_y \cdot (\nabla_y \chi A) = \nabla_y \cdot A, \quad y \in \mathbb{T}^d.$$ 

Then, for any $\epsilon > 0$ and $K \subset PD_{\alpha,\beta}$ compact in $L^2(\mathbb{T}^d; \mathbb{R}^{d \times d})$, there exists an FNO $\Psi(\cdot; \theta^*) : K \to \dot{H}^1(\mathbb{T}^d; \mathbb{R}^d)$ such that

$$\sup_{A \in K} \|\Psi^d(A) - \Psi(A; \theta^*)\|_{\dot{H}^1} < \epsilon.$$
\[ K \subset \text{BV}(\mathbb{T}^d; \mathbb{R}^{d \times d}) \cap PD_{\alpha,\beta} \subset L^2(\mathbb{T}^d; \mathbb{R}^{d \times d}). \]
Learning Error for Voronoi Microstructure

(a) True $\chi_1$

(b) Error $\chi_1$

(c) True $\nabla \chi_1$

(d) Error $\nabla \chi_1$
Test Error versus Data Size

The graph shows the relationship between the relative H1 error and the number of training data points. The error decreases as the number of training data points increases. The dashed line represents the theoretical behavior: $\frac{1}{\sqrt{N}}$. The solid line with triangles represents the Voronoi method, and the solid line with circles represents the smooth method.
Test Error versus Model Size

(a) Smooth

Relative H1 Error vs. Number of Fourier Modes

(b) Voronoi

Relative H1 Error vs. Number of Fourier Modes

- Modes*Width = 144
- Modes*Width = 288
- Modes*Width = 576
- Modes*Width = 1152
Talk Outline

Set-Up

Algorithms

Analysis

Application I: Material-Dependent Constitutive Models

Application II: History-Dependent Constitutive Models

Conclusions
### Big Picture

#### Multiscale Problem

*Pavliotis and AMS ’08 [20][Ch12]*

Displacement $u^\varepsilon(x, t)$, stress $\sigma^\varepsilon(x, t)$, $0 < \varepsilon \ll 1$. $F=MA$:

$$\rho \partial_t^2 u^\varepsilon = \nabla \cdot (\sigma^\varepsilon) + f, \quad \sigma^\varepsilon = \Psi^\varepsilon \left( \{\nabla u^\varepsilon\}, \frac{x}{\varepsilon} \right).$$

#### Homogenized Problem

*Bensoussan, Lions, Papanicolaou [1]*

Approximate $u^\varepsilon = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \ldots$

Determine map $\Psi$, so that small scales are removed in $u_0$. $F=MA$:

$$\rho \partial_t^2 u_0 = \nabla \cdot (\sigma) + f, \quad \sigma = \Psi(\{\nabla u_0\}).$$

#### Operator Learning

- $\mathcal{U}$ histories of strain $\{\nabla u_0\}$; $\mathcal{V}$ histories of stress $\{\sigma\}$.
- Approximate $\Psi \approx \Psi_{NN}$
Viscoelasticity I  Francfort and Suquet ’86 [6]

Quasi-Static Viscoelasticity Multiscale Problem

\[-\nabla \cdot (\sigma^\epsilon) = f, \quad x \in \Omega\]

\[\sigma^\epsilon = \nu^\epsilon \partial_t \nabla u^\epsilon + E^\epsilon \nabla u^\epsilon\]

\[E^\epsilon(x) = E \left(\frac{x}{\epsilon}\right), \quad \nu^\epsilon(x) = \nu \left(\frac{x}{\epsilon}\right), \quad E, \nu : \mathbb{T}^d \rightarrow \mathbb{R}.\]

Laplace Transform, Homogenize, Invert

\[-\nabla \cdot (\bar{\sigma}^\epsilon) = f, \quad x \in \Omega\]

\[\bar{\sigma}^\epsilon = (s \nu^\epsilon + E^\epsilon) \nabla \bar{u}^\epsilon.\]

▶ Introduces memory.
Theorem (Piecewise-Constant Homogenization: Memory)

In piecewise-constant case and in dimension $d = 1$ homogenized equation for $u_0$ is Markovian:

\[-\nabla \cdot (\sigma) = f, \quad x \in \Omega,\]
\[
\sigma = \nu' \partial_t \nabla u_0 + E' \nabla u_0 + \langle 1, r \rangle
\]
\[
\partial_t r_\ell = -\alpha_\ell r_\ell + \beta_\ell \nabla u_0, \quad \ell \in \{1, 2, \cdots, L\},
\]

for some choice of $E' \in \mathbb{R}_+$, $\nu' \in \mathbb{R}_+$, $\alpha \in \mathbb{R}_+^L$, $\beta \in \mathbb{R}^L$, $L \in \mathbb{Z}_+$.

- Homogenization introduces memory.
- In $d = 1$ (approximate) Markovian structure.
- Dimension $d > 1$ ?
Viscoelasticity III: Operator Learning

**True Solution Map**

Let $\Psi : \mathcal{U} \rightarrow \mathcal{V}$ be the map such that the homogenized constitutive relation is

$$\sigma = \Psi(\{\nabla u_0\}).$$

**Goal: Supervised Learning (RNO-NET)**

Learn map $\Psi_{RNO} : \mathcal{U} \rightarrow \mathcal{V}$ approximating $\Psi$ with the form

$$\sigma = F(\nabla u_0, \partial_t \nabla u_0, r)$$

$$\partial_t r = G(r, \nabla u_0), \quad r(0) = 0.$$

- RNO – Recurrent neural operator.
- Dimension of memory variable $r$ has to be learned.
Figure: Viscoelasticity: trained model performs well on test samples
Viscoelasticity V:
Choosing the Number of Hidden Variables

Figure: Absolute $L^2$ error of RNNs trained with different numbers of hidden variables on different piecewise-constant viscoelastic materials.
Viscoelasticity VI: Time Discretization Invariance

Figure: Viscoelasticity models trained with and without access to the strain rate variable: preferred model exhibits more invariance to time discretization of test trajectories
Talk Outline

Set-Up

Algorithms

Analysis

Application I: Material-Dependent Constitutive Models

Application II: History-Dependent Constitutive Models

Conclusions
Conclusions

1. Algorithms:
   - define on function space;
   - then learn;
   - leads to models which transfer between discretizations.

2. Analysis:
   - universal approximation theory well-developed;
   - complexity (cost versus error) incompletely understood;
   - what solution is found via optimization?

3. Applications:
   - cheap surrogates;
   - scientific discovery;
   - constitutive laws.
References


References II

On universal approximation and error bounds for Fourier neural operators.

*Machine Learning and Scientific Computing*.

Neural operator: Learning maps between function spaces with applications to pdes.

Operator learning with pca-net: upper and lower complexity bounds.

The nonlocal neural operator: Universal approximation.

Error estimates for deeponets: a deep learning framework in infinite dimensions.

The curse of dimensionality in operator learning.

Fourier neural operator for parametric partial differential equations.
References III


Plasticity Multiscale Problem

\[
\rho \partial_t^2 u^\epsilon = \nabla \cdot \sigma^\epsilon + f, \quad x \in \Omega
\]
\[
\partial_t \xi^\epsilon = K(\xi^\epsilon, \nabla u^\epsilon), \quad x \in \Omega
\]
\[
\sigma^\epsilon = \Psi^\epsilon \left( \nabla u^\epsilon, \xi^\epsilon, \frac{x}{\epsilon} \right)
\]
Plasticity II

Homogenized Plasticity Problem

\[ \rho \partial_t^2 u_0 = \nabla \cdot \sigma_0 + f, \quad x \in \Omega \]

\[ \sigma_0 = \Psi(\{\nabla u_0\}) \]
Plasticity III: Operator Learning

**Goal: Supervised Learning (PCA-NET)**

Learn map $\Psi_{PCA} : U \rightarrow V$ approximating $\Psi$. In particular causality must be learned.

**Goal: Supervised Learning (RNO-NET)**

Learn map $\Psi_{RNO} : U \rightarrow V$ approximating $\Psi$ with the form

$$\sigma = F(\nabla u_0, \partial_t \nabla u_0, r)$$

$$\partial_t r = G(r, \nabla u_0), \quad r(0) = 0.$$
Figure: Viscoplasticity: Trained model performs well on test samples. Left: Input strains. Right: Output truth and approximation.
Plasticity V:
Choosing the Number of Hidden Variables

Figure: Viscoplasticity: 3D polycrystal (different hidden variable counts)
Figure: Models with viscoplastic (VP) and elasto-viscoplastic (E-VP) architecture trained on data from an E-VP material: preferred model exhibits more time discretization invariance.
Learning Error for Varying Microstructures (I)

Smooth Microstructure

Star Inclusion Microstructure
Learning Error for Varying Microstructures (II)

Square Inclusion Microstructure

Voronoi Microstructure

(a) True $\chi_1$

(b) Error $\chi_1$

(c) True $\nabla \chi_1$

(d) Error $\nabla \chi_1$