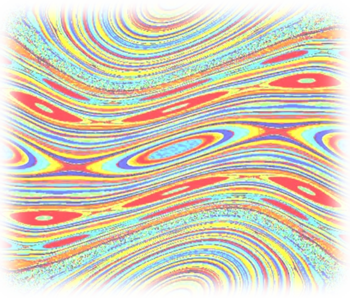
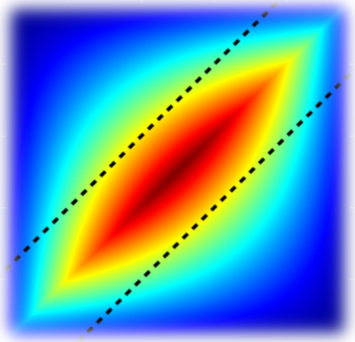




An Overview of Operator Learning

Infinite-dimensional NLA

Nicolas Boullé and Matthew Colbrook

University of Cambridge

13/05/2024

B., Townsend *"A Mathematical Guide to Operator Learning,"* **Handbook of Numerical Analysis**, 2024.

C., *"The Multiverse of Dynamic Mode Decomposition Algorithms,"* **Handbook of Numerical Analysis**, 2024.

Shameless plug: Nicolas' *SIAG/LA Best Paper Prize: Recovering Green's Functions with Randomized Numerical Linear Algebra* - **5pm Thursday**

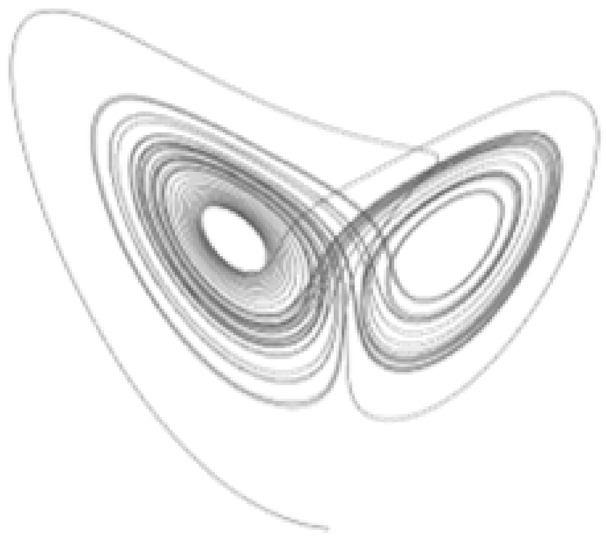
Part 1: Data-driven dynamical systems

State $x \in \Omega \subseteq \mathbb{R}^d$.

Unknown function $F: \Omega \to \Omega$ governs dynamics: $x_{n+1} = F(x_n)$.

Goal: Learning from data $\{x^{(m)}, y^{(m)} = F(x^{(m)})\}_{m=1}^{M}$.

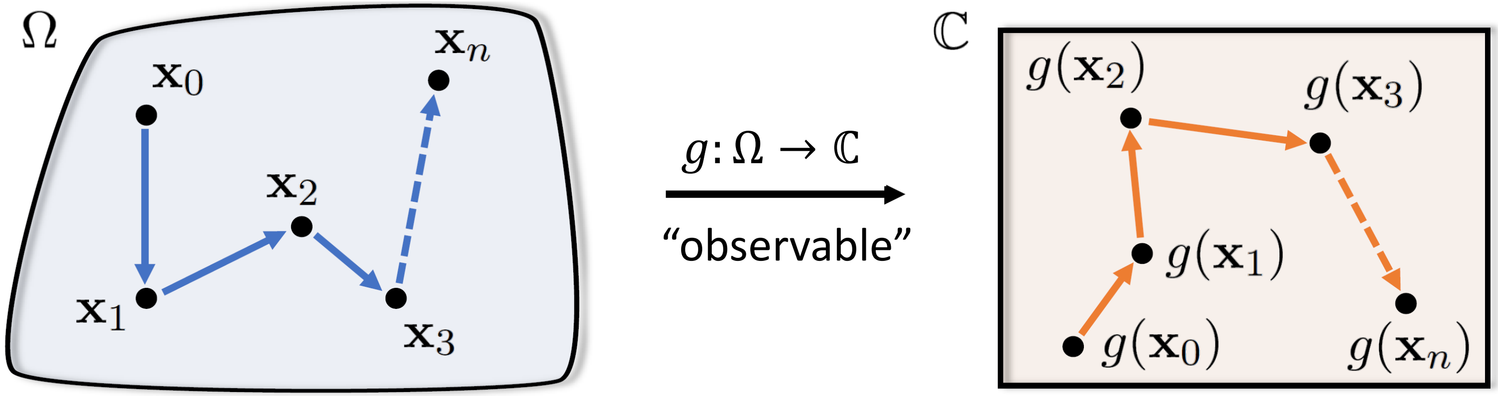
Applications: chemistry, climatology, control, electronics, epidemiology, finance, fluids, molecular dynamics, neuroscience, plasmas, robotics, video processing, etc.



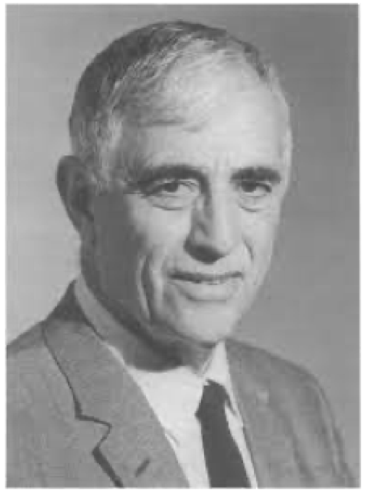
Surveys:

- Brunton, Budišić, Kaiser, Kutz, *"Modern Koopman theory for dynamical systems,"* SIAM Review, 2022.
- Budišić, Mohr, Mezić, *"Applied Koopmanism,"* Chaos, 2012.
- C., *"The Multiverse of Dynamic Mode Decomposition Algorithms,"* Handbook of Numerical Analysis, 2024.

Koopman Operator $\mathcal{K}$: A global linearization



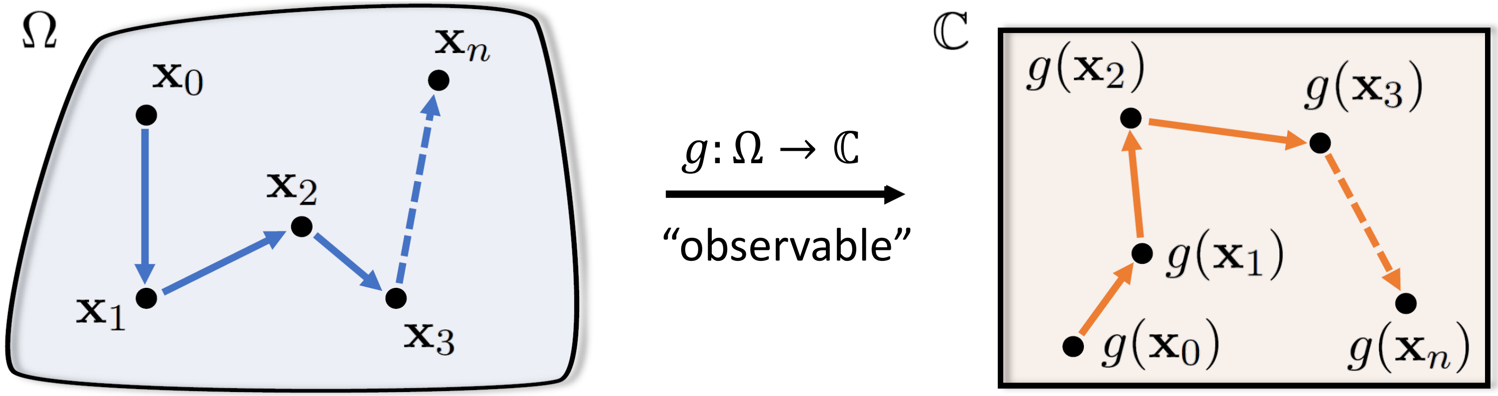
Koopman



von Neumann

- Koopman, *"Hamiltonian systems and transformation in Hilbert space,"* **Proc. Natl. Acad. Sci. USA**, 1931.
- Koopman, v. Neumann, *"Dynamical systems of continuous spectra,"* **Proc. Natl. Acad. Sci. USA**, 1932.

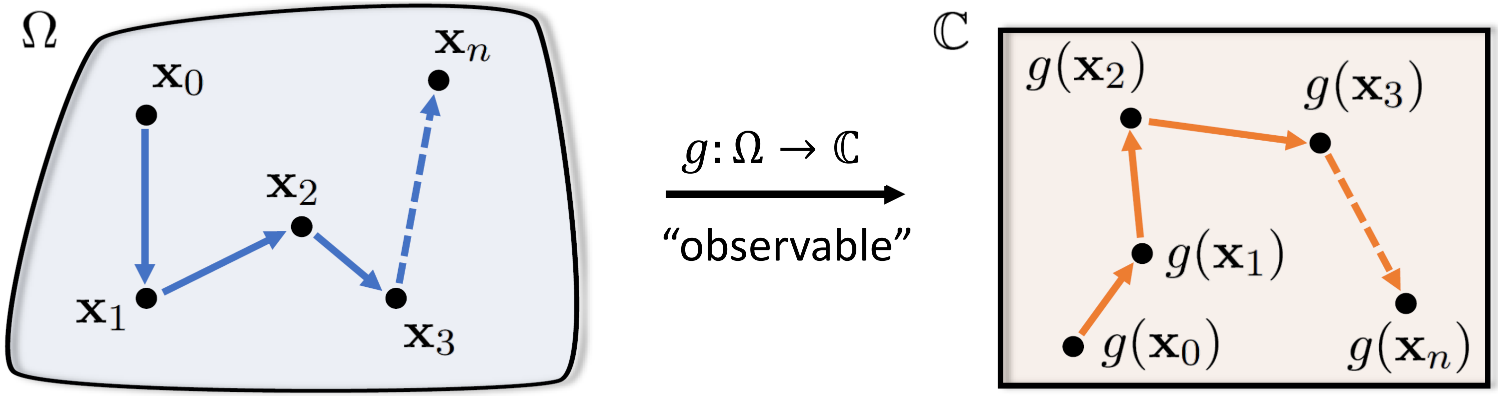
Koopman Operator $\mathcal{K}$: A global linearization



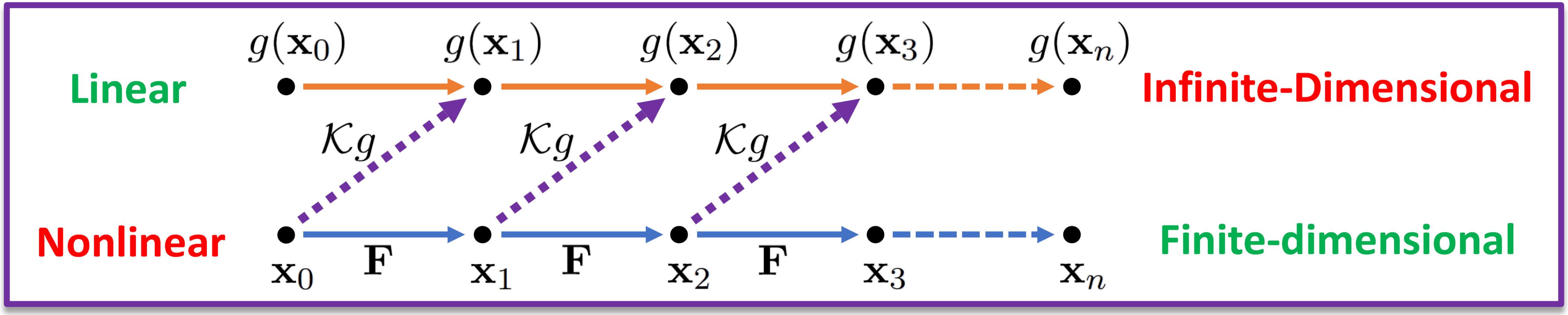
- $\mathcal{K}$ acts on functions $g: \Omega \to \mathbb{C}$, $[\mathcal{K}g](x) = g(F(x))$.
- Function space: $g \in L^2(\Omega, \omega)$, positive measure ω, inner product $\langle \cdot, \cdot \rangle$.

- Koopman, *"Hamiltonian systems and transformation in Hilbert space,"* **Proc. Natl. Acad. Sci. USA**, 1931.
- Koopman, v. Neumann, *"Dynamical systems of continuous spectra,"* **Proc. Natl. Acad. Sci. USA**, 1932.

Koopman Operator $\mathcal{K}$: A global linearization

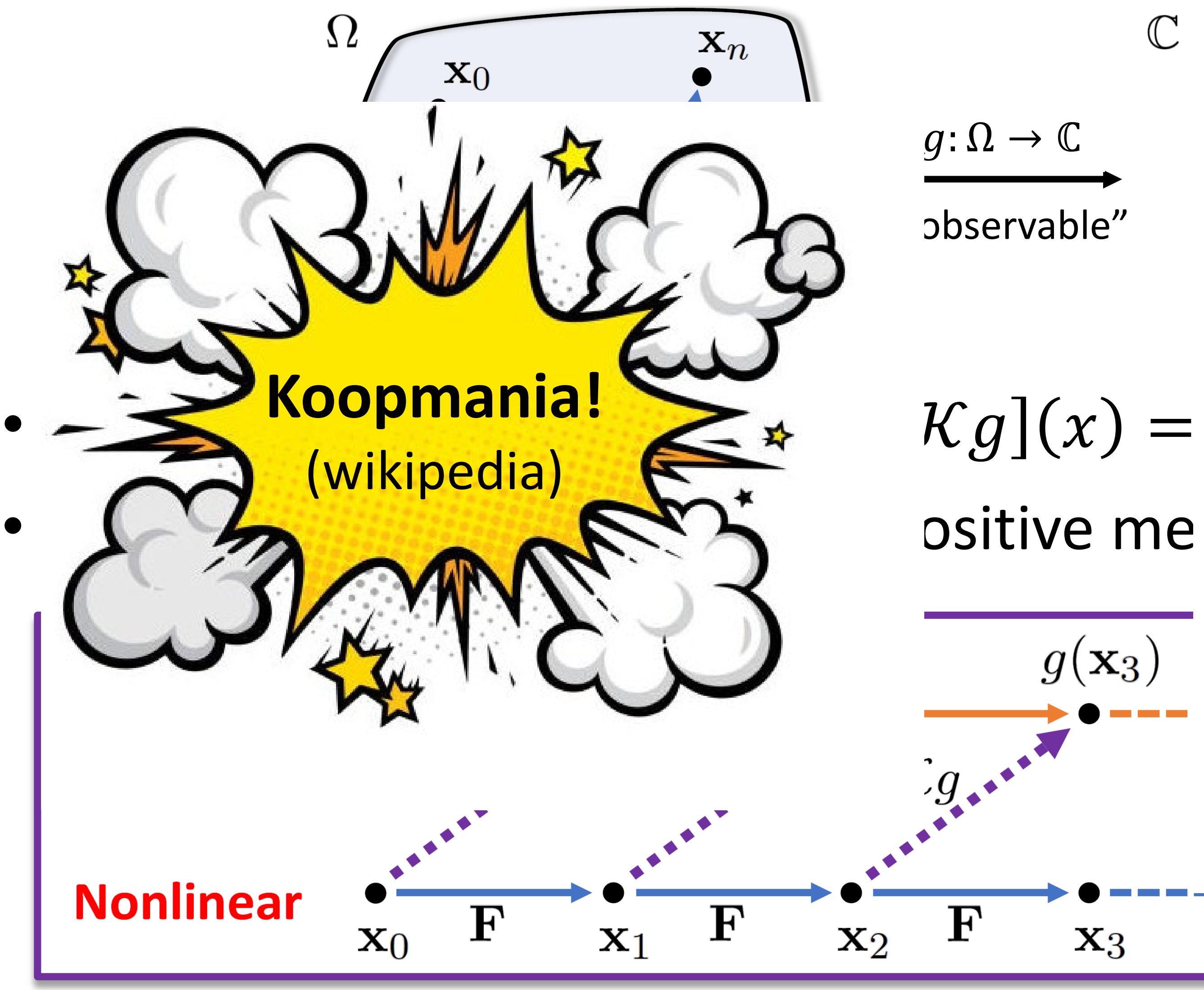


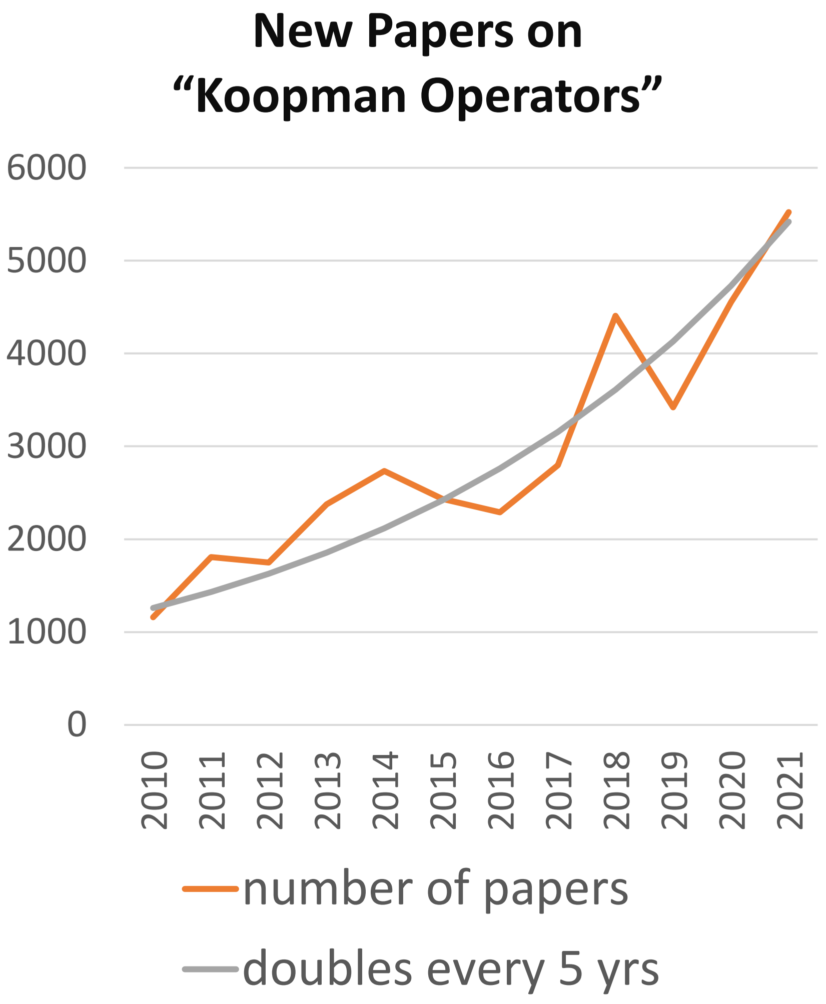
- $\mathcal{K}$ acts on functions $g: \Omega \to \mathbb{C}$, $[\mathcal{K}g](x) = g(F(x))$.
- Function space: $g \in L^2(\Omega, \omega)$, positive measure ω, inner product $\langle \cdot, \cdot \rangle$.



- Koopman, *"Hamiltonian systems and transformation in Hilbert space,"* **Proc. Natl. Acad. Sci. USA**, 1931.
- Koopman, v. Neumann, *"Dynamical systems of continuous spectra,"* **Proc. Natl. Acad. Sci. USA**, 1932.

Koopman Operator $\mathcal{K}$: A global linearization





- Koopman, “*Hamiltonian systems and transformation in Hilbert space,*” **Proc. Natl. Acad. Sci. USA**, 1931.
- Koopman, v. Neumann, “*Dynamical systems of continuous spectra,*” **Proc. Natl. Acad. Sci. USA**, 1932.

Koopman mode decomposition

$x_{n+1} = F(x_n)$

$[\mathcal{K}g](x) = g(F(x))$

eigenfunction of $\mathcal{K}$

generalized eigenfunction of $\mathcal{K}$

$$g(x) = \sum_{\text{eigenvalues } \lambda_j} c_{\lambda_j} \varphi_{\lambda_j}(x) + \int_{-\pi}^{\pi} \phi_{\theta,g}(x) \, \mathrm{d}\theta$$

$$g(x_n) = [\mathcal{K}^n g](x_0) = \sum_{\text{eigenvalues } \lambda_j} c_{\lambda_j} {\lambda_j}^n \varphi_{\lambda_j}(x_0) + \int_{-\pi}^{\pi} e^{in\theta} \phi_{\theta,g}(x_0) \, \mathrm{d}\theta$$

Encodes: geometric features, invariant measures, transient behavior, long-time behavior, coherent structures, quasiperiodicity, etc.

GOAL: Data-driven approximation of $\mathcal{K}$ and its spectral properties.

- Mezić, "*Spectral properties of dynamical systems, model reduction and decompositions,*" **Nonlinear Dynamics**, 2005.

Koopman mode decomposition

$x_{n+1} = F(x_n)$

$[\mathcal{K}g](x) = g(F(x))$

eigenfunction of $\mathcal{K}$

generalized eigenfunction of $\mathcal{K}$

$$g(x) = \sum_{\text{eigenvalues } \lambda_j} c_{\lambda_j} \varphi_{\lambda_j}(x) + \int_{-\pi}^{\pi} \phi_{\theta,g}(x)\, \mathrm{d}\theta$$

$$g(x_n) = [\mathcal{K}^n g](x_0) = \sum_{\text{eigenvalues } \lambda_j} c_{\lambda_j} \lambda_j{}^n \varphi_{\lambda_j}(x_0) + \int_{-\pi}^{\pi} e^{in\theta} \phi_{\theta,g}(x_0)\, \mathrm{d}\theta$$

Encodes: geometric features, invariant measures, transient behavior, long-time behavior, coherent structures, quasiperiodicity, etc.

GOAL: Data-driven approximation of $\mathcal{K}$ and its spectral properties.

- Mezić, "*Spectral properties of dynamical systems, model reduction and decompositions*," **Nonlinear Dynamics**, 2005.

Example: Dynamic Mode Decomposition (DMD)

Given dictionary $\{\psi_1, \dots, \psi_N\}$ of functions $\psi_j: \Omega \to \mathbb{C}$,

$$\{x^{(m)}, y^{(m)} = F(x^{(m)})\}_{m=1}^{M}$$

$$\langle \psi_k, \psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(x^{(m)})} \psi_k(x^{(m)}) = \Bigg[\underbrace{\begin{pmatrix} \psi_1(x^{(1)}) & \cdots & \psi_N(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(x^{(M)}) & \cdots & \psi_N(x^{(M)}) \end{pmatrix}^*}_{\Psi_X} \underbrace{\begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_M \end{pmatrix}}_{W} \underbrace{\begin{pmatrix} \psi_1(x^{(1)}) & \cdots & \psi_N(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(x^{(M)}) & \cdots & \psi_N(x^{(M)}) \end{pmatrix}}_{\Psi_X} \Bigg]_{jk}$$

$$\langle \mathcal{K}\psi_k, \psi_j \rangle \approx \sum_{m=1}^{M} w_m \overline{\psi_j(x^{(m)})} \underbrace{\psi_k(y^{(m)})}_{[\mathcal{K}\psi_k](x^{(m)})} = \Bigg[\underbrace{\begin{pmatrix} \psi_1(x^{(1)}) & \cdots & \psi_N(x^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(x^{(M)}) & \cdots & \psi_N(x^{(M)}) \end{pmatrix}^*}_{\Psi_X} \underbrace{\begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_M \end{pmatrix}}_{W} \underbrace{\begin{pmatrix} \psi_1(y^{(1)}) & \cdots & \psi_N(y^{(1)}) \\ \vdots & \ddots & \vdots \\ \psi_1(y^{(M)}) & \cdots & \psi_N(y^{(M)}) \end{pmatrix}}_{\Psi_Y} \Bigg]_{jk}$$

$$\mathcal{K} \longrightarrow \mathbb{K} = (\Psi_X{}^* W \Psi_X)^{-1} \Psi_X{}^* W \Psi_Y = (\sqrt{W}\Psi_X)^{\dagger} \sqrt{W} \Psi_Y \in \mathbb{C}^{N \times N}$$

- Schmid, "*Dynamic mode decomposition of numerical and experimental data,*" **J. Fluid Mech.**, 2010.
- Rowley, Mezić, Bagheri, Schlatter, Henningson, "*Spectral analysis of nonlinear flows,*" **J. Fluid Mech.**, 2009.
- Kutz, Brunton, Brunton, Proctor, "*Dynamic mode decomposition: data-driven modeling of complex systems,*" **SIAM**, 2016.
- Williams, Kevrekidis, Rowley "*A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition,*" **J. Nonlinear Sci.**, 2015.

Nonlinear system

↓ “Hit it with Koopman!”
Data becomes quadrature

A ***linear*** algebra problem

Catch: Infinite dimensions

(almost always necessary to deal with nonlinearity)

A simple example on $\ell^2(\mathbb{Z})$
(Why it's not just DMD + MATLAB's "eig")

$$\begin{pmatrix} \ddots & \ddots & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & 0 & 1 & \\ & & & & 0 & \ddots \\ & & & & & \ddots \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} \in \mathbb{C}^{N\times N}$$

Two-way infinite

- Spectrum is unit circle.
- Spectrum is stable.
- Continuous spectra.
- Unitary evolution.

- Spectrum is $\{0\}$.
- Spectrum is unstable.
- Discrete spectra.
- Nilpotent evolution.

Example might look silly, but lots of Koopman operators are built up from operators like these!

Challenges of infinite dimensions

$$\mathrm{Sp}(\mathcal{K}) = \{\lambda \in \mathbb{C} : \mathcal{K} - \lambda I \text{ is not invertible}\}$$

Caution

- **Too much:** Spurious eigenvalues $\lambda \notin \mathrm{Sp}(\mathcal{K})$
- **Too little:** Miss parts of $\mathrm{Sp}(\mathcal{K})$
- **Continuous spectra** ($\mathrm{Sp}(\mathcal{K})$ not just eigenvalues!)
- **Verification.**
- **Preserving key structures.**
- **Choice of dictionary.**
- **Instability** (non-normal $\mathcal{K}$, non-normal discretizations of normal $\mathcal{K}$)
- **Compact representations** (low-rank, neural networks, pseudoeigenfunctions etc.)

Truncate/discretize

$$\mathcal{K} \longrightarrow \mathbb{K} \in \mathbb{C}^{N\times N}$$

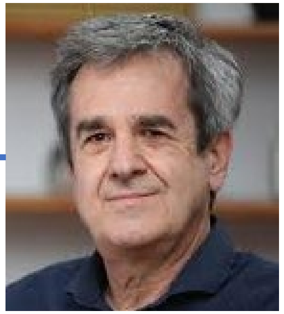
- C., Townsend, *"Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems,"* **Commun. Pure Appl. Math.**, 2024.

Today’s talks

- **Claire Valva:** *On Consistent Spectral Approximation of Koopman Operators Using Resolvent Compactification*
- **Dimitrios Giannakis:** *Learning Dynamical Systems with the Spectral Exterior Calculus*
- **Nathan Kutz:** *Shallow Recurrent Decoders for Encoding Operators*
- **Jean-Christophe Loiseau:** *Low-Rank Approximation of the Koopman Operator*
- **Zlatko Drmač:** *A Data Driven Koopman-Schur Decomposition for Computational Analysis of Nonlinear Dynamics*
- **Igor Mezić:** *Operator Is the Model*
- **Catherine Drysdale:** *Rigged DMD: Data-Driven Koopman Decompositions via Generalized Eigenfunctions*



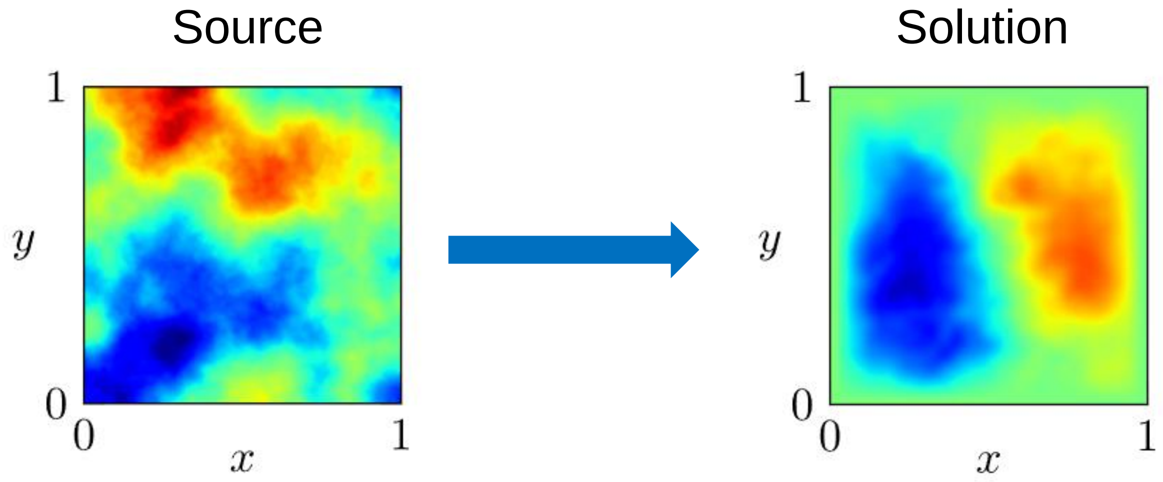




Operator learning for PDEs

Unknown partial differential operator governing the system $\mathcal{L}(u) = f$

Goal: Learn the solution operator from input-output pairs $\{(f_j, u_j)\}$



Many applications:

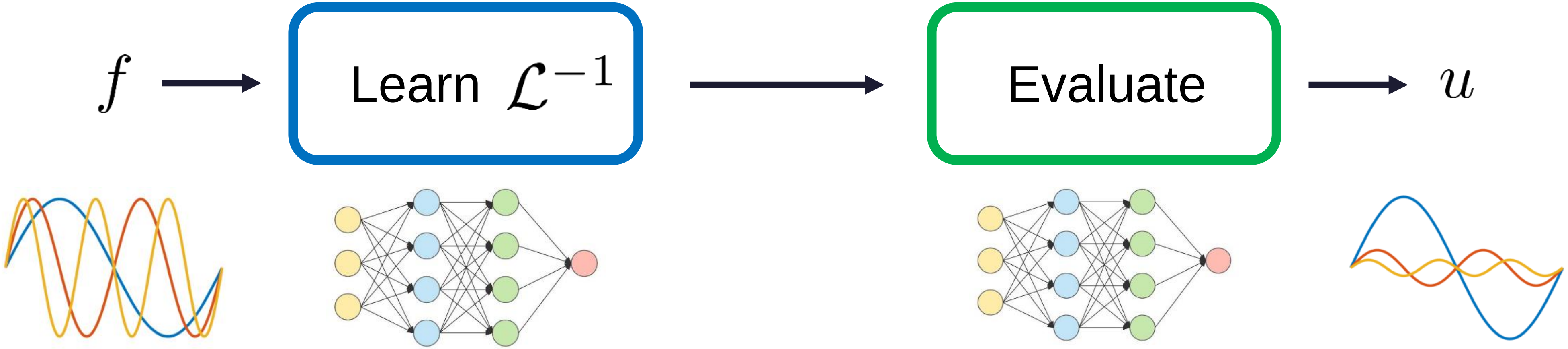
Fluid dynamics, continuum mechanics, quantum mechanics, weather forecasting, reduced-order modelling, parameter optimization, …

Recent surveys:

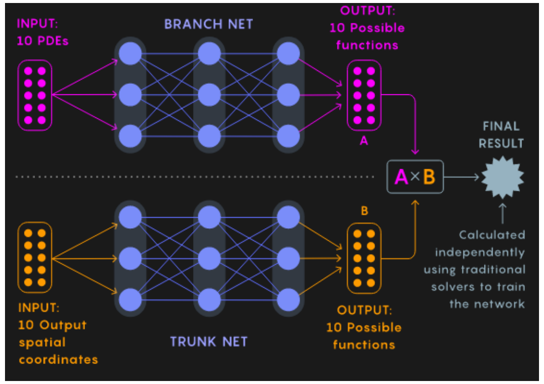
- B., Townsend, ''A Mathematical Guide to Operator Learning'', 2023.
- Kovachki, Lanthaler, Stuart, ''Operator Learning: Algorithms and Analysis'', 2023.

Neural operators

<u>Aim:</u> Approximate solution operators of unknown PDEs $\mathcal{L}(u) = f$

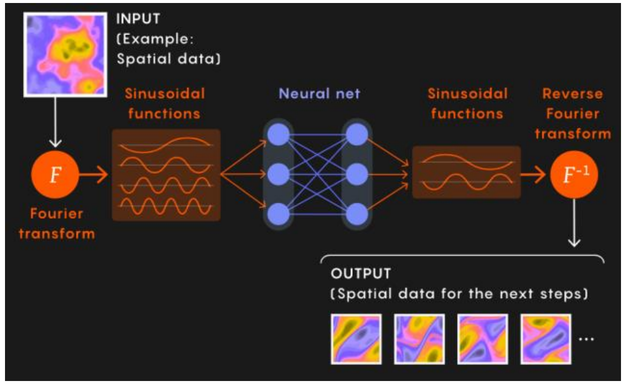


DeepONet



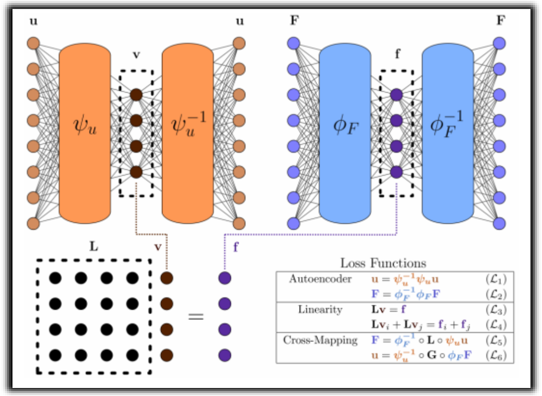
[Quanta Magazine; Lu et al, 2021]

Fourier Neural Operator



[Quanta Magazine; Li et al, 2020]

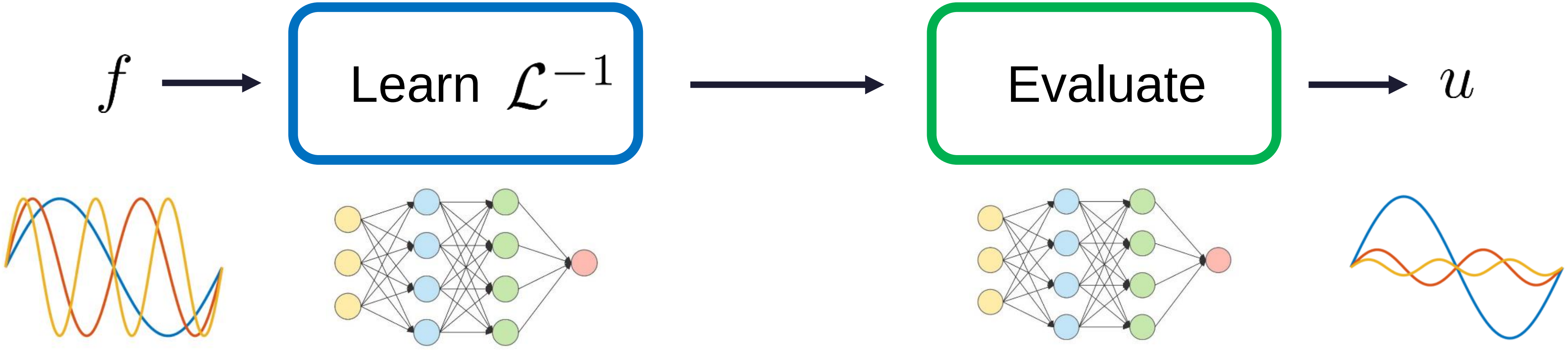
DeepGreen



[Gin et al., 2020]

Neural operators

<u>Aim:</u> Approximate solution operators of unknown PDEs $\mathcal{L}(u) = f$



Neural operators can be defined as compositions of integral operators and nonlinear activation functions:

$$u_{i+1}(x) = \sigma\left(\int_{\Omega_i} K^{(i)}(x,y)u_i(y)\,\mathrm{d}y + b_i(x)\right)$$

Connections with NLA and main challenges

Neural operators can be defined as compositions of integral operators and nonlinear activation functions:

$$u_{i+1}(x) = \sigma\left(\int_{\Omega_i} K^{(i)}(x,y)u_i(y)\,\mathrm{d}y + b_i(x)\right)$$

Numerical linear algebra insights ⟶ Choice of K

Connections with NLA and main challenges

Class of PDE → Solution operator property → Matrix structure

Operator learning motivates numerical linear algebra research on matrix recovery problems and infinite-dimensional NLA.

Challenges:

1. How should we discretize neural operators?
2. How much training data is required?
3. What PDE can we learn?
4. Which neural operator architecture should we use?

Discretization

How should we discretize neural operators?

Representation Equivalent Neural Operators: a Framework for Alias-free Operator Learning

Francesca Bartolucci[1] **Emmanuel de Bézenac**[2] **Bogdan Raonić**[2,3]

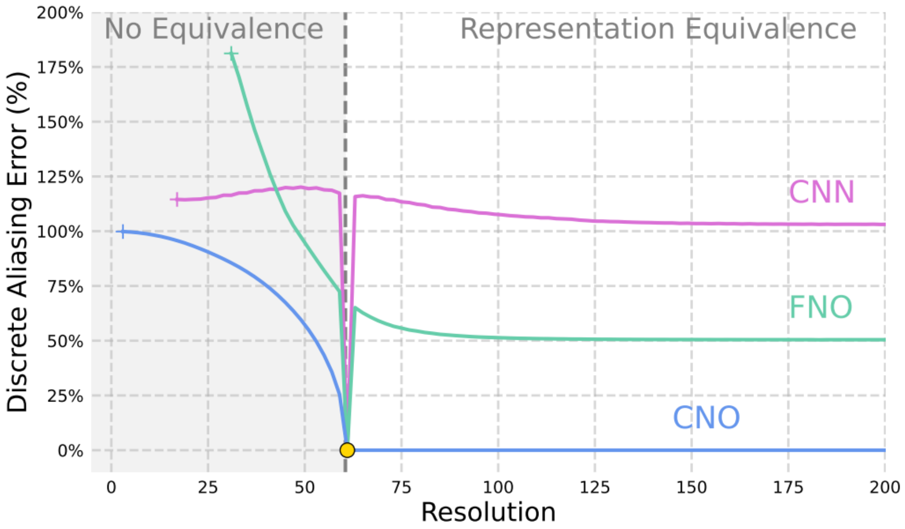
Roberto Molinaro[2] **Siddhartha Mishra**[2,3] **Rima Alaifari**[2,3]

1 Delft University of Technology, Netherlands
2 Seminar for Applied Mathematics, ETH, Zurich, Switzerland
3 ETH AI Center, Zurich, Switzerland

Abstract

Recently, *operator learning*, or learning mappings between infinite-dimensional function spaces, has garnered significant attention, notably in relation to learning partial differential equations from data. Conceptually clear when outlined on paper, neural operators necessitate discretization in the transition to computer implementations. This step can compromise their integrity, often causing them to deviate from the underlying operators. This research offers a fresh take on neural operators with a framework *Representation equivalent Neural Operators (ReNO)* designed to address these issues. At its core is the concept of operator aliasing, which measures inconsistency between neural operators and their discrete representations. We explore this for widely-used operator learning techniques. Our findings detail how aliasing introduces errors when handling different discretizations and grids and loss of crucial continuous structures. More generally, this framework not only sheds light on existing challenges but, given its constructive and broad nature, also potentially offers tools for developing new neural operators.

Aliasing-free operator learning

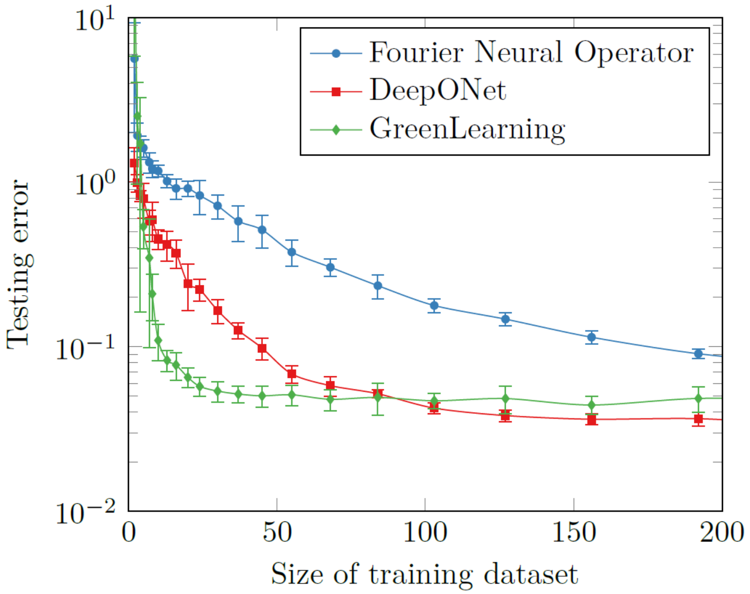


Francesca Bartolucci

Sample complexity

How much data do we need to train neural operators?

Diana Halikias



Data-efficient operator learning

PNAS BRIEF REPORT | APPLIED MATHEMATICS

Elliptic PDE learning is provably data-efficient

Nicolas Boullé[a,1], Diana Halikias[b], and Alex Townsend[b]

Edited by David Donoho, Stanford University, Stanford, CA; received March 8, 2023; accepted July 21, 2023

Partial differential equations (PDE) learning is an emerging field that combines physics and machine learning to recover unknown physical systems from experimental data. While deep learning models traditionally require copious amounts of training data, recent PDE learning techniques achieve spectacular results with limited data availability. Still, these results are empirical. Our work provides theoretical guarantees on the number of input–output training pairs required in PDE learning. Specifically, we exploit randomized numerical linear algebra and PDE theory to derive a provably data-efficient algorithm that recovers solution operators of three-dimensional uniformly elliptic PDEs from input–output data and achieves an exponential convergence rate of the error with respect to the size of the training dataset with an exceptionally high probability of success.

Learning feasibility

What PDE can we learn?

Operator learning without the adjoint

Nicolas Boullé NB690@CAM.AC.UK
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Cambridge, CB3 0WA, UK

Diana Halikias DH736@CORNELL.EDU
Department of Mathematics
Cornell University
Ithaca, NY 14853, USA

Samuel E. Otto OTTONCDR@UW.EDU
AI Institute in Dynamic Systems
University of Washington
Seattle, WA 98103, USA

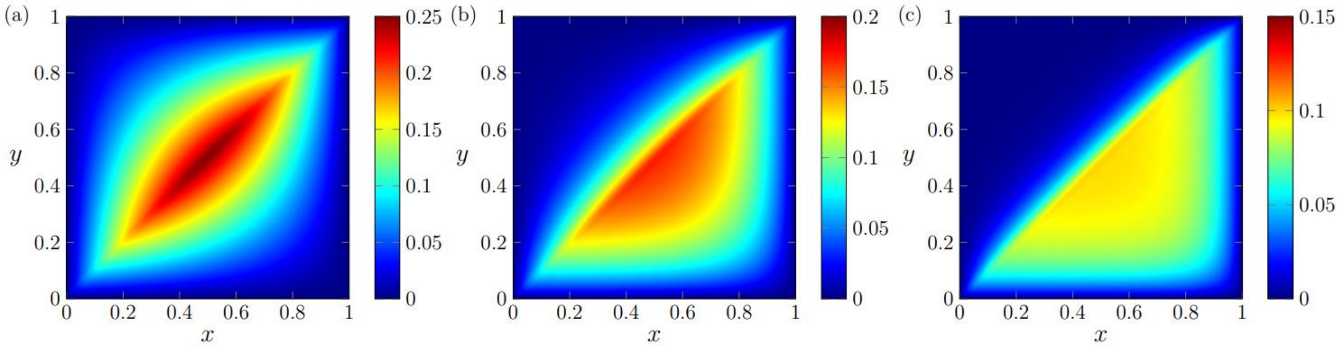
Alex Townsend TOWNSEND@CORNELL.EDU
Department of Mathematics
Cornell University
Ithaca, NY 14853, USA

Abstract

There is a mystery at the heart of operator learning: how can one recover a non-self-adjoint operator from data without probing the adjoint? Current practical approaches suggest that one can accurately recover an operator while only using data generated by the forward action of the operator without access to the adjoint. However, naively, it seems essential to sample the action of the adjoint. In this paper, we partially explain this mystery by proving that without querying the adjoint, one can approximate a family of non-self-adjoint infinite-dimensional compact operators via projection onto a Fourier basis. We then apply the result to recovering Green's functions of elliptic partial differential operators and derive an adjoint-free sample complexity bound. While existing theory justifies low sample complexity in operator learning, ours is the first adjoint-free analysis that attempts to close the gap between theory and practice.

Samuel Otto

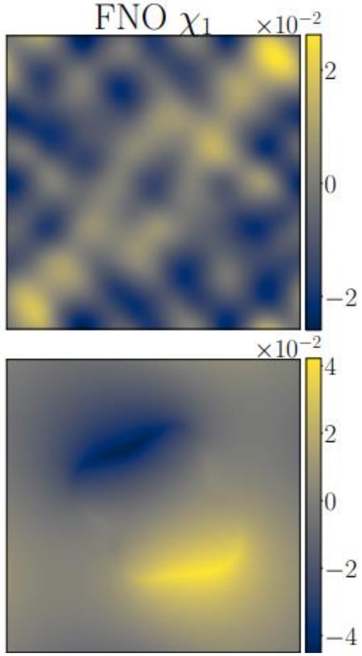
Adjoint-free operator learning



Architectures and approximation theory

Which neural operator architecture should we use?

Margaret Trautner



Multiscale operator learning

LEARNING HOMOGENIZATION FOR ELLIPTIC OPERATORS*

KAUSHIK BHATTACHARYA†, NIKOLA B. KOVACHKI‡, AAKILA RAJAN†, ANDREW M. STUART†, AND MARGARET TRAUTNER†

Abstract.
Multiscale partial differential equations (PDEs) arise in various applications, and several schemes have been developed to solve them efficiently. Homogenization theory is a powerful methodology that eliminates the small-scale dependence, resulting in simplified equations that are computationally tractable while accurately predicting the macroscopic response. In the field of continuum mechanics, homogenization is crucial for deriving constitutive laws that incorporate microscale physics in order to formulate balance laws for the macroscopic quantities of interest. However, obtaining homogenized constitutive laws is often challenging as they do not in general have an analytic form and can exhibit phenomena not present on the microscale. In response, data-driven learning of the constitutive law has been proposed as appropriate for this task. However, a major challenge in data-driven learning approaches for this problem has remained unexplored: the impact of discontinuities and corner interfaces in the underlying material. These discontinuities in the coefficients affect the smoothness of the solutions of the underlying equations. Given the prevalence of discontinuous materials in continuum mechanics applications, it is important to address the challenge of learning in this context; in particular, to develop underpinning theory that establishes the reliability of data-driven methods in this scientific domain. The paper addresses this unexplored challenge by investigating the learnability of homogenized constitutive laws for elliptic operators in the presence of such complexities. Approximation theory is presented, and numerical experiments are performed which validate the theory in the context of learning the solution operator defined by the cell problem arising in homogenization for elliptic PDEs.

Summary

- 9:45-10:05 : Francesca Bartolucci

Representation Equivalent Neural Operators: a Framework for Alias-free Operator Learning

- 11:00-11:20 : Diana Halikias

Elliptic PDE Learning Is Provably Data-Efficient

- 15:45-16:05 : Samuel Otto

Operator learning without the adjoint

- 16:10-16:30 : Margaret Trautner

Operator Learning for Multiscale Elliptic PDEs with History Dependence