Data-efficient Kernel Methods for Learning Differential Equations and their Solution Operators

Bamdad Hosseini
Department of Applied Mathematics
University of Washington

Data Assimilation and Inverse Problems for Digital Twins Workshop
Institute for Mathematical and Statistical Innovation

Oct 8, 2025

New talk ...
new typos ...

Matthew Thorpe

Acknowledgements

Yasamin Jalalian, Juan Felipe Osorio Ramirez, Alexander Hsu, Bamdad Hosseini, and Houman Owhadi, **Data-Efficient Kernel Methods for Learning Differential Equations and Their Solution Operators: Algorithms and Error Analysis**, arXiv:2503.01036, 2025

- AH, JFOP, and BH were supported by
  - NSF-DMS-2208535, "Machine Learning for Bayesian Inverse Problems"
  - NSF-DMS-2337678, "CAREER: Gaussian Processes for Scientific Machine Learning: Theoretical Analysis and Computational Algorithms"
- YJ and HO were supported by
  - NSF-GEO-2425908, "CAIG: Discovering the Law of Stress Transfer and Earthquake Dynamics in a Fault Network using a Computational Graph Discovery Approach"
  - FA9550-20-1-0358, "Machine Learning and Physics-Based Modeling and Simulation"
  - FOA-AFRL-AFOSR-2023-0004, "Mathematics of Digital Twins"
  - DE-SC0023163, "SEA-CROGS: Scalable, Efficient, and Accelerated Causal Reasoning Operators, Graphs and Spikes for Earth and Embedded Systems"
  - Jet Propulsion Laboratory PDRDF 24AW0133

Outline

Overview

Solving PDEs / Learning PDEs / Operator Learning

Generic nonlinear PDE

$$\mathfrak{P}(u) = f$$

- (PDE solvers) Given $\mathfrak{P}, f$ find $\widehat{u}$ such that $\mathfrak{P}(\widehat{u}) \approx f$

$$\widehat{\mathfrak{P}^{-1}} : f \mapsto u$$

- (Learning PDEs) Given training data $\{u_m, f_m\}_{m=1}^M$ find $\widehat{\mathfrak{P}}$ such that $\widehat{\mathfrak{P}}(u) \approx f$

$$\widehat{\mathfrak{P}} : u \mapsto f$$

- (Operator learning) Given training data $\{u_m, f_m\}_{m=1}^M$ find $\widehat{\mathfrak{P}^{-1}}$ such that $\widehat{\mathfrak{P}^{-1}}(f) \approx u$

$$\widehat{\mathfrak{P}^{-1}} : f \mapsto u$$

[1] Brunton, Proctor, and Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems".

[2] Kovachki, Lanthaler, and Stuart, "Operator learning: Algorithms and analysis".

- Sparse observation meshes $X_m := \{x_{m,1}, \ldots, x_{m,N}\} \subset \Omega$

Goal

Given sparse and noisy observations

$$\{u_m(X_m) + \epsilon_m, f_m\}_{m=1}^M$$

learn $\widehat{\mathfrak{P}} \approx \mathfrak{P}$ and $\widehat{\mathfrak{P}^{-1}} \approx \mathfrak{P}^{-1}$.

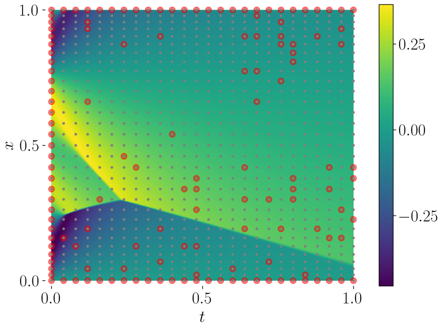
Idea

First learn $\widehat{\mathfrak{P}}$ assuming it is local. Then "invert" $\widehat{\mathfrak{P}}$ to obtain $\widehat{\mathfrak{P}^{-1}}$.

E.g. Burgers' PDE

$$\mathfrak{P}(u) = u_t + uu_x - \nu u_{xx} = 0$$

While $\mathfrak{P}$ is simple, the solution map $\mathfrak{P}^{-1} : u(x,0) \mapsto u(x,t)$ is complex.



Problem Setup

$$\mathfrak{P}(u)(x) = f(x) \qquad x \in \Omega$$

$$\begin{aligned} \mathfrak{P}(u)(x) &= p \circ (x, Du(x), D^2u(x), \dots,) \\ &= p \circ \Phi(u, x) \end{aligned}$$

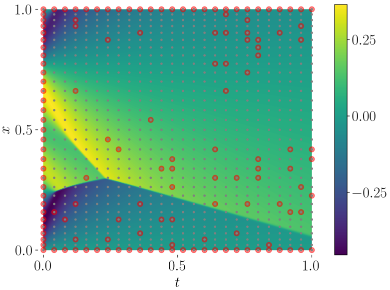
- **(known)** Mapping $\Phi : \mathcal{U} \times \Omega \to \mathbb{R}^Q$, linear in u and nonlinear in x
- **(unknown)** Nonlinear function $p : \mathbb{R}^Q \to \mathbb{R}$

E.g. Nonlinear, Variable Coefficient, Elliptic PDE

$$\begin{aligned} \mathfrak{P}(u) &= -\partial_x \left(a(x)\partial_x u\right) + \alpha u^3 = f \\ \Phi(u, x) &= (x, u(x), u_x(x), u_{xx}(x)) \\ p(s_1, s_2, s_3, s_4) &= -a_x(s_1)s_3 - a(s_1)s_4 + \alpha s_2^3 \end{aligned}$$

Kernel Equation Learning (KEqL): Formulation

- Assumption $\mathfrak{P} = p \circ \Phi(u, x) = f(x)$
- Training data $(u_m(X_m), f_m)_{m=1}^{M}$
- Banach spaces $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$ and $(\mathcal{P}, \|\cdot\|_{\mathcal{P}})$
- Optimal recovery problem



$$(\widehat{u}_m, \widehat{p}) = \underset{v_m \in \mathcal{U}, q \in \mathcal{P}}{\arg\min} \|q\|_{\mathcal{P}} + \lambda \sum_{m=1}^{M} \|v_m\|_{\mathcal{U}}$$

s.t. $v_m(X_m) = u_m(X_m)$ (Observation/data)

$q \circ \Phi(v_m, x) = f_m(x), \quad \forall x \in \Omega$ (PDE constraint)

Reproducing Kernel Hilbert Spaces (RKHS)

- Set Ω
- Positive definite and symmetric (PDS) kernel $K : \Omega \times \Omega \to \mathbb{R}$:
  - $K(x, x') = K(x', x), \ \forall x, x' \in \Omega$
  - For any set of points $X = \{x_1, \ldots, x_N\} \subset \Omega$ the matrix $K(X, X) = (K(x_i, x_j))$ is PDS.
- Pre-Hilbert space

$$\mathcal{K}_0 := \left\{ f : \Omega \to \mathbb{R} \middle| f = \sum_{j=1}^{N_f} \alpha_{f,j} K(\cdot, x_{f,j}) = K(\cdot, X_f)\boldsymbol{\alpha}_f \right\}$$

- Define inner product

$$\langle f, g \rangle_{\mathcal{K}_0} := \boldsymbol{\alpha}_f^T K(X_f, X_g)\boldsymbol{\alpha}_g$$

- Complete $\mathcal{K}_0$ to get RKHS $\mathcal{K}$
- $\mathcal{K}$ is uniquely defined by K
- Reproducing property: $\forall f \in \mathcal{K}$

$$f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{K}}$$

- In fact, for any $\phi \in \mathcal{K}^\star$

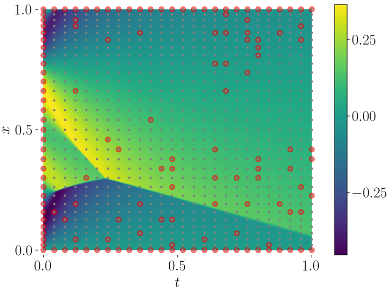
$$\phi(f) = \langle f, K(\cdot, \phi) \rangle_{\mathcal{K}}$$

where

$$K(x, \phi) = \phi(K(x, \cdot))$$

Kernel Equation Learning (KEqL): Formulation

- Assumption $\mathfrak{P} = p \circ \Phi(u, x) = f(x)$
- Training data $(u_m(X_m), f_m)_{m=1}^M$
- RKHS space $\mathcal{U}$ with kernel U
- RKHS space $\mathcal{P}$ with kernel P
- Collocation mesh $X \supseteq \cup_m X_m$



$$(\widehat{u}_m, \widehat{p}) = \underset{v_m \in \mathcal{U}, q \in \mathcal{P}}{\arg\min} \|q\|_{\mathcal{P}}^2 + \lambda \sum_{m=1}^{M} \|v_m\|_{\mathcal{U}}^2$$

s.t. $v_m(X_m) = u_m(X_m)$ (Observation/data)

$q \circ \Phi(v_m, X) = f_m(X)$ (Discrete PDE constraint)

KEqL: All-at-once Inversion

$$
\begin{aligned}
(\widehat{u}_m, \widehat{p}) = \underset{v_m \in \mathcal{U}, q \in \mathcal{P}}{\arg\min} \; & \|q\|_{\mathcal{P}}^2 + \lambda \sum_{m=1}^{M} \|v_m\|_{\mathcal{U}}^2 \\
\text{s.t.} \qquad & v_m(X_m) = u_m(X_m) \qquad && \text{(Observation/data)} \\
& q \circ \Phi(v_m, X) = f_m(X) && \text{(Discrete PDE constraint)}
\end{aligned}
$$

[3] Chen, Liu, and Sun, "Physics-informed learning of governing equations from scarce data".
[4] Sun, Liu, and Sun, "Physics-informed Spline Learning for Nonlinear Dynamics Discovery".
[5] Haber and Oldenburg, "Joint inversion: a structural approach".
[6] Kaltenbacher, "Regularization based on all-at-once formulations for inverse problems".

KEqL: Algorithm Summary

- Relax equality constraints

$$(\widehat{u}_m, \widehat{p}) = \underset{v_m \in \mathcal{U}, q \in \mathcal{P}}{\arg\min} \|q\|_{\mathcal{P}}^2 + \sum_{m=1}^{M} \lambda \|v_m\|_{\mathcal{U}}^2 + \lambda' \|u_m(X^m) - v_m(X^m)\|_2^2 + \lambda'' \|q \circ \Phi(v_m, X) - f_m(X)\|_2^2$$

- Apply kernel trick (representer theorem) or use feature maps
- Solve using a Levenberg–Marquardt-type algorithm

Operator Learning via KEqL

- Given KEqL solution $\widehat{p}$ approximate PDE solution map $\mathfrak{P}^{-1}$ with pseudo inverse of

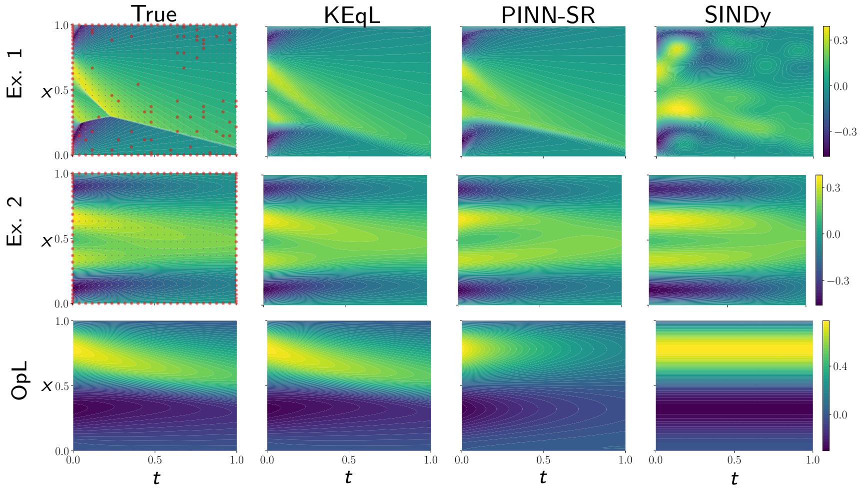
$$\widehat{\mathfrak{P}} := \widehat{p} \circ \Phi.$$

$$\widehat{\mathfrak{P}}^{\dagger}(f) := \operatorname*{arg\,min}_{v \in \mathcal{U}} \|v\|_{\mathcal{U}}^2 + \lambda'' \|\widehat{p} \circ \Phi(v, X) - f(X)\|_2^2.$$

[7] Chen et al., "Solving and learning nonlinear PDEs with Gaussian processes".
[8] Long et al., "A kernel framework for learning differential equations and their solution operators".

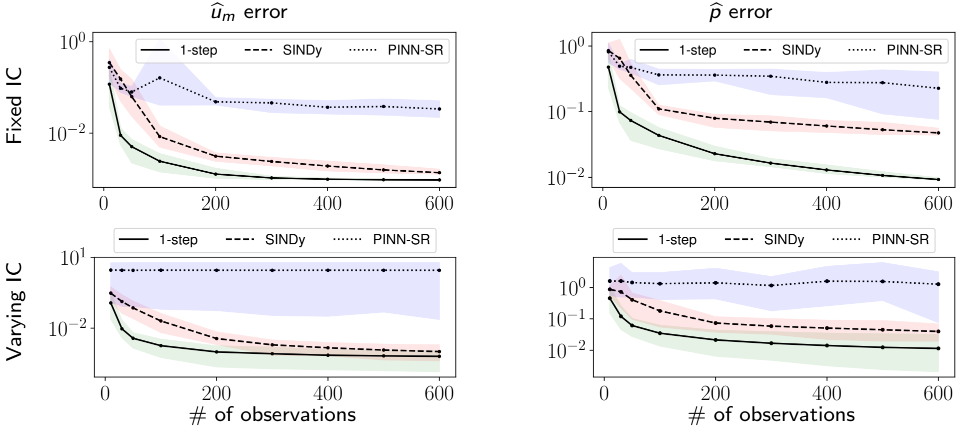
Burgers' Benchmark: Shocks with Sparse Data

$$\mathfrak{P}(u) = u_t + uu_x - 0.1u_x x = 0 \quad \text{plus B.C and I.C}$$



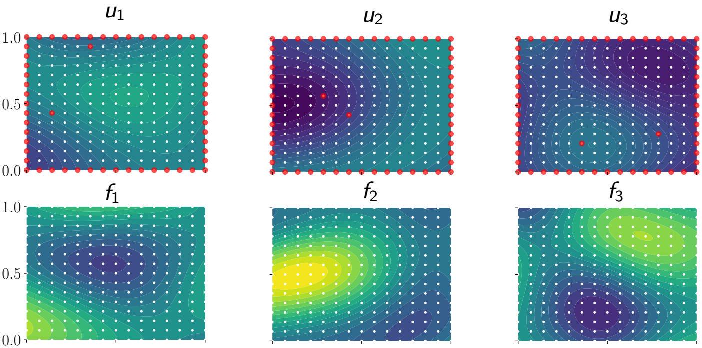
Burgers' Benchmark: Comparison to PINN-SR and SINDy

- Large performance gap with PINN model



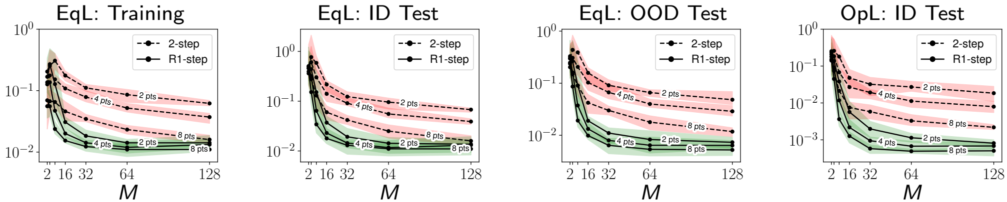
Darcy Flow PDE: Variable Coefficients with Sparse Data

$$\mathfrak{P}(u) = -\text{div } a \nabla u = f \qquad a(x) = \exp(\sin(\cos(x_1) + \cos(x_2))) \qquad + \text{B.C.}$$



Darcy Flow PDE: EqL and OpL Errors

- M – # of solution and source pairs $(u_m(X), f_m)_{m=1}^{M}$
- 2-step – Methods like SINDy that infer u_m first before learning p
- R1-step – KEqL with reduced basis for better performance



Quantitative Error Analysis

Theorem [JOHHO] (Part 1)

Consider $\{u_m(X), f_m\}_{m=1}^M$ defined on a smooth domain $\Omega \subset \mathbb{R}^d$ and observation mesh $X_{\text{obs}} \subset \Omega$ with $|X_{\text{obs}}| = N$. Suppose that $u^m \in H^\gamma(\Omega)$ for $\gamma > d/2 +$ order of $\mathfrak{P}$ and $\mathfrak{P} = p \circ \Phi$ such that $\Phi(u,x) \in \mathbb{R}^Q$. Define the fill distance

$$\rho := \sup_{x' \in \Omega} \inf_{x \in X_{\text{obs}}} |x - x'|,$$

Then under sufficient smoothness assumptions it holds for $0 \leq \gamma' \leq \gamma$

$$\sum_{m=1}^M \|u_m - \widehat{u}_m\|^2_{H^{\gamma'}(\Omega)} \lesssim \rho^{2(\gamma-\gamma')} \left(\|p\|^2_{\mathcal{P}} + \sum_{m=1}^M \|u_m\|^2_{\mathcal{U}} \right).$$

[9] Zhang and Schaeffer, "On the convergence of the SINDy algorithm".
[10] Scholl et al., "The uniqueness problem of physical law learning".
[11] He, Zhao, and Zhong, "How much can one learn a partial differential equation from its solution?"
[12] Boullé, Halikias, and Townsend, "Elliptic PDE learning is provably data-efficient".

Quantitative Error Analysis

Theorem [JOHHO] (Part 2)

Consider $\{u_m(X), f_m\}_{m=1}^M$ defined on a smooth domain $\Omega \subset \mathbb{R}^d$ and observation mesh $X_{\mathrm{obs}} \subset \Omega$ with $|X_{\mathrm{obs}}| = N$. Suppose that $u^m \in H^\gamma(\Omega)$ for $\gamma > d/2 + \text{order of } \mathfrak{P}$ and $\mathfrak{P} = p \circ \Phi$ such that $\Phi(u, x) \in \mathbb{R}^Q$ and $p \in H^\eta(\mathbb{R}^Q)$ for $\eta > Q/2$. Define the set $S := \cup_{m=1}^M \cup_{x \in X_{\mathrm{obs}}} \Phi(u_m, x)$ along with the fill distances

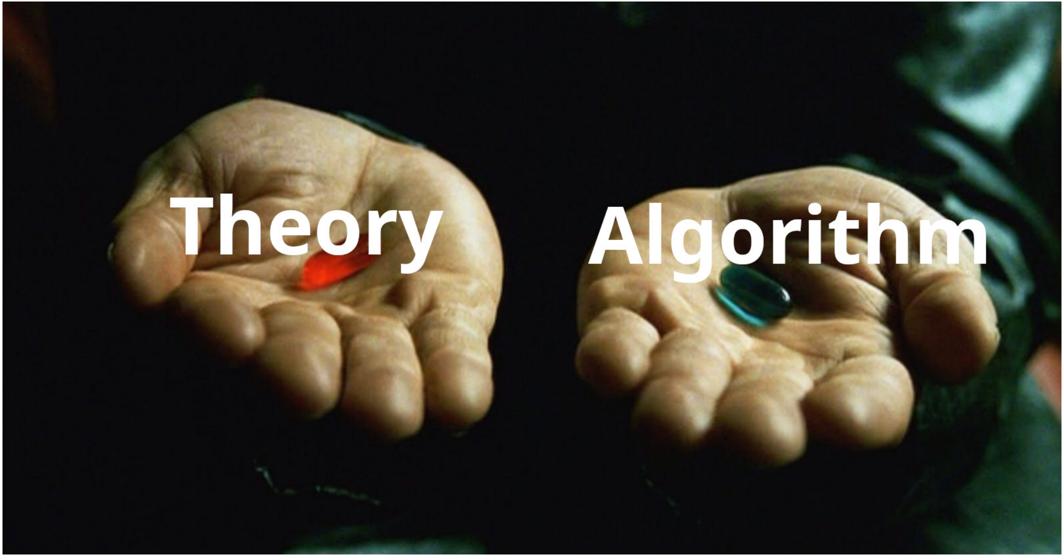
$$\rho := \sup_{x' \in \Omega} \inf_{x \in X_{\mathrm{obs}}} |x - x'|, \qquad \varrho(B) := \sup_{s' \in B} \inf_{s \in S \cap B} |s - s'|,$$

for any smooth and bounded set $B \subset \mathbb{R}^Q$. Then under sufficient smoothness assumptions it holds for $d/2 + \text{order of } \mathfrak{P} \leq \gamma' \leq \gamma$

$$\sum_{m=1}^M \|p - \widehat{p}\|_{L^\infty(B)} \lesssim \left[\rho^{\gamma-\gamma'} + \varrho(B)^{\eta - Q/2}\right] \left(\|p\|_{\mathcal{P}} + \sum_{m=1}^M \|u_m\|_{\mathcal{U}}\right).$$

Summary

- A kernel method for learning PDEs and filtering solutions
- Joint/simultaneous recovery of solutions and PDE form
- Successful recovery of PDEs and their solution maps with very scarce data
- Optimization problem is amenable to quasi-newton algorithms, easier and more efficient training
- Better performance compared to some neural net models or two step learning
- Operator learning through equation learning, going well beyond typical operator learning setups
- Quantitative convergence analysis, reminiscent of Sobolev sampling inequalities for scattered data approximation


Theory
Algorithm

Outline

Algorithms

RKHS Representer Theorems

$$\widehat{f} = \arg\min_{f \in \mathcal{K}} \|f\|_{\mathcal{K}} \quad \text{s.t.} \quad \boldsymbol{\phi}(f) = \mathbf{z}$$

- Shorthand notation $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N) \in (\mathcal{K}^\star)^N$
- Data $\mathbf{z} \in \mathbb{R}^N$
- Closed form solution

$$\widehat{f} = K(\cdot, \boldsymbol{\phi}) K(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}, \qquad \|\widehat{f}\|_{\mathcal{K}}^2 = \mathbf{z}^T K(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}$$

- Vector field $K(\cdot, \boldsymbol{\phi}) = (K(\cdot, \phi_1), \ldots, K(\cdot, \phi_N)) \in \mathcal{K}^N$
- Matrix $K(\boldsymbol{\phi}, \boldsymbol{\phi}) \in \mathbb{R}^{N \times N}$

$$K(\boldsymbol{\phi}, \boldsymbol{\phi})_{ij} = \phi_i(K(\cdot, \phi_j))$$

Feature Map Perspective

$$\widehat{f} = \arg\min_{f \in \mathcal{K}} \|f\|_{\mathcal{K}} \quad \text{s.t.} \quad \boldsymbol{\phi}(f) = \mathbf{z}$$

- Solution

$$\widehat{f} = K(\cdot, \boldsymbol{\phi}) K(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}, \qquad \|\widehat{f}\|_{\mathcal{K}}^2 = \mathbf{z}^T K(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}$$

- Coefficient vector $\boldsymbol{\alpha} = K(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}$

$$\widehat{f} = K(\cdot, \boldsymbol{\phi}) \boldsymbol{\alpha} = \sum_{j=1}^{N} \alpha_j K(\cdot, \phi_j)$$

- Using reproducing property

$$\|\widehat{f}\|_{\mathcal{K}}^2 = \boldsymbol{\alpha}^T K(\boldsymbol{\phi}, \boldsymbol{\phi}) \boldsymbol{\alpha}$$

Reformulate KEqL

$$(\widehat{u}_m, \widehat{p}) = \underset{v_m \in \mathcal{U}, q \in \mathcal{P}}{\arg\min} \|q\|_{\mathcal{P}}^2 + \sum_{m=1}^{M} \|v_m\|_{\mathcal{U}}^2 + \|u_m(X^m) - v_m(X^m)\|_2^2 + \|q \circ \Phi(v_m, X) - f_m(X)\|_2^2$$

- Take $\lambda = \lambda' = \lambda'' = 1$ for simplicity
- Reduced model $v_m = U(\cdot, X)\boldsymbol{\alpha}_m$ and define $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M)$
- Recall X is the collocation mesh
- Define $S(\boldsymbol{\alpha}) = \cup_{m=1}^{M} \cup_{x \in X} \Phi(K(\cdot, X)\boldsymbol{\alpha}, x)$ and $S(\boldsymbol{\alpha}_m) = \cup_{x \in X} \Phi(K(\cdot, X)\boldsymbol{\alpha}_m, x)$
- Take $q = P(\cdot, S(\boldsymbol{\alpha}))\beta$

$$(\widehat{\boldsymbol{\alpha}}, \widehat{\beta}) = \underset{\boldsymbol{\alpha}, \beta}{\arg\min} \, \beta^T P(S(\boldsymbol{\alpha}), S(\boldsymbol{\alpha}))\beta + \sum_{m=1}^{M} \boldsymbol{\alpha}_m^T U(X, X)\boldsymbol{\alpha}_m + \|u_m(X^m) - U(X^m, X)\boldsymbol{\alpha}_m\|_2^2 + \|P(S(\boldsymbol{\alpha}_m), S(\boldsymbol{\alpha}))\beta - f_m(X)\|_2^2$$

Reformulate KEqL

$$
\begin{aligned}
(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\arg\min}\, & \boldsymbol{\beta}^T P(S(\boldsymbol{\alpha}), S(\boldsymbol{\alpha}))\boldsymbol{\beta} + \sum_{m=1}^{M} \boldsymbol{\alpha}_m^T U(X, X)\boldsymbol{\alpha}_m \\
& + \|u_m(X^m) - U(X^m, X)\boldsymbol{\alpha}_m\|_2^2 + \|P(S(\boldsymbol{\alpha}_m), S(\boldsymbol{\alpha}))\boldsymbol{\beta} - f_m(X)\|_2^2
\end{aligned}
$$

- Compositional structure
- Quadratic in $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\beta}$
- Quadratic in $\boldsymbol{\beta}$ for fixed $\boldsymbol{\alpha}$
- Propose sequential quadratic approximations
- Additional regularization to control deviation in each step, similar to Levenberg–Marquardt (LM)

LM for KEqL

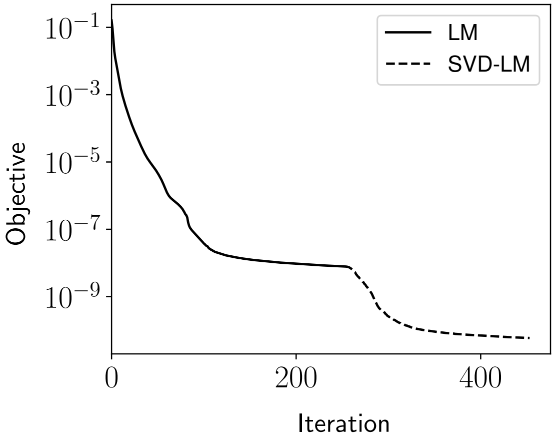
$$(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) = \operatorname*{arg\,min}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\beta}^T P(S(\boldsymbol{\alpha}), S(\boldsymbol{\alpha}))\boldsymbol{\beta} + \sum_{m=1}^{M} \boldsymbol{\alpha}_m^T U(X,X)\boldsymbol{\alpha}_m$$
$$+ \|u_m(X^m) - U(X^m, X)\boldsymbol{\alpha}_m\|_2^2 + \|P(S(\boldsymbol{\alpha}_m), S(\boldsymbol{\alpha}))\boldsymbol{\beta} - f_m(X)\|_2^2$$

Minimizing sequence $(\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)})$ given by

$$(\widehat{\boldsymbol{\alpha}}^{(k+1)}, \widehat{\boldsymbol{\beta}}^{(k+1)}) := \operatorname*{arg\,min}_{\boldsymbol{\alpha},\boldsymbol{\beta}} \boldsymbol{\beta}^T P(S(\boldsymbol{\alpha}^{(k)}), S(\boldsymbol{\alpha}^{(k)}))\boldsymbol{\beta} + \sum_{m=1}^{M} \boldsymbol{\alpha}_m^T U(X,X)\boldsymbol{\alpha}_m$$
$$+ \|u_m(X^m) - U(X^m, X)\boldsymbol{\alpha}_m\|_2^2$$
$$+ \left\| \left[P(S(\boldsymbol{\alpha}_m^{(k)}), S(\boldsymbol{\alpha}^{(k)})) + \nabla_{\boldsymbol{\alpha}} P(S(\boldsymbol{\alpha}_m^{(k)}), S(\boldsymbol{\alpha}^{(k)}))(\boldsymbol{\alpha} - \boldsymbol{\alpha}^{(k)}) \right] \boldsymbol{\beta} - f_m(X) \right\|_2^2$$
$$+ \sigma_k(\boldsymbol{\alpha} - \boldsymbol{\alpha}^{(k)})^T U(X,X)(\boldsymbol{\alpha} - \boldsymbol{\alpha}^{(k)}) + \sigma_k'(\boldsymbol{\beta} - \boldsymbol{\beta}^{(k)})^T P(S(\boldsymbol{\alpha}_m^{(k)}), S(\boldsymbol{\alpha}^{(k)}))(\boldsymbol{\beta} - \boldsymbol{\beta}^{(k)})$$

Choose σ_k, σ_k' based on decrease of objective, established heuristics for LM

LM for KEqL



Theory

Sobolev Sampling Inequalities

Sobolev Sampling Inequality

Suppose $\Omega \subset \mathbb{R}^d$ is a bounded set with Lipschitz boundary and consider a set of points $X = \{x_1, \ldots, x_N\} \subset \overline{\Omega}$ with fill distance $h_X := sup_{x\in\Omega} \inf_{x'\in X} \|x - x'\|_2$. Let $u|_X$ denote the restriction of u to the set X. Further consider $\gamma > d/2$ and $0 \le \gamma' \le \gamma$ and let $u \in H^\gamma(\Omega)$.

1. (Noiseless) Suppose $u|_X = 0$. Then there exists $h_0 > 0$ so that whenever $h_X \le h_0$ we have $\|u\|_{H^{\gamma'}(\Omega)} \le C_\Omega h_X^{\gamma-\gamma'} \|u\|_{H^\gamma(\Omega)}$, where $C_\Omega > 0$ is a constant that depends only on Ω.
2. (Noisy) Suppose $u|_X \neq 0$. Then there exists $h_0 > 0$ so that whenever $h_X \le h_0$ we have $\|u\|_{L^\infty(\Omega)} \le C_\Omega h_X^{\gamma-d/2} \|u\|_{H^\gamma(\Omega)} + 2\|u|_X\|_\infty,$ where $C_\Omega > 0$ is a constant that depends only on Ω.

Controlling the Filtering Error

$$(\widehat{u}_m, \widehat{p}) = \underset{v_m \in \mathcal{U}, q \in \mathcal{P}}{\arg\min} \|q\|_{\mathcal{P}}^2 + \sum_{m=1}^{M} \|v_m\|_{\mathcal{U}}^2$$

$$\text{s.t.} \qquad v_m(X_{\text{obs}}) = u_m(X_{\text{obs}}) \qquad \text{(Observation/data)}$$

$$q \circ \Phi(v_m, X_{\text{obs}}) = f_m(X_{\text{obs}}) \quad \text{(Discrete PDE constraint)}$$

- Apply Sobolev sampling inequality, recalling $\rho = \sup_{x' \in \Omega} \inf_{x \in X_{\text{obs}}} |x - x'|$

$$\|\widehat{u}_m - u_m\|_{H^{\gamma'}(\Omega)} \lesssim \rho^{(\gamma - \gamma')} \|\widehat{u}_m - u_m\|_{H^{\gamma}(\Omega)}$$

- Sum up, use assumed embedding $\mathcal{U} \subset H^{\gamma}(\Omega)$, and optimality

$$\sum_m \|\widehat{u}_m - u_m\|_{H^{\gamma'}(\Omega)}^2 \lesssim \rho^{2(\gamma - \gamma')} \left(\sum_m \|\widehat{u}_m\|_{H^{\gamma}(\Omega)}^2 + \|\widehat{u}_m\|_{H^{\gamma}(\Omega)}^2 \right)$$

$$\lesssim \rho^{2(\gamma - \gamma')} \left(\sum_m \|\widehat{u}_m\|_{\mathcal{U}}^2 + \|u_m\|_{\mathcal{U}}^2 \right) \lesssim \rho^{2(\gamma - \gamma')} \left(\|p\|_{\mathcal{P}}^2 + \sum_m \|u_m\|_{\mathcal{U}}^2 \right)$$

Controlling the Equation Error

$$(\widehat{u}_m, \widehat{p}) = \underset{v_m \in \mathcal{U}, q \in \mathcal{P}}{\arg\min} \ \|q\|_{\mathcal{P}}^2 + \sum_{m=1}^{M} \|v_m\|_{\mathcal{U}}^2$$

$$\text{s.t.} \qquad v_m(X_{\text{obs}}) = u_m(X_{\text{obs}})$$

$$q \circ \Phi(v_m, X_{\text{obs}}) = f_m(X_{\text{obs}}) = p \circ \Phi(u_m, X_{\text{obs}})$$

- Observe this is "almost" an interpolation problem for p
- $S = \cup_m \cup_{x \in X_{\text{obs}}} \Phi(u_m, x)$ then constraint is

$$q(S + \text{noise}) = p(S)$$

- Akin to total least squares!
- We need to do more work

Controlling the Equation Error

- Basic idea

$$\begin{aligned} \widehat{p}(S+\delta S) &= p(S) \\ \widehat{p}(S) + \nabla \widehat{p}(S)\delta S &\approx p(S) \\ \Rightarrow |\widehat{p}(S) - p(S)| &\approx |\nabla \widehat{p}(S)\delta S| \end{aligned}$$

- Use Sobolev sampling inequality again but with noisy RHS
- Local Lipschitz assumption for $q \in \mathcal{P}$:

$$|q(s) - q(s')| \leq C(B)\|q\|_{\mathcal{P}}|s - s'|, \quad \forall s, s' \in B,$$

for any bounded set $B \subset \mathbb{R}^Q$

Controlling the Equation Error

$$\widehat{p} \circ \Phi(\widehat{u}_m, x_k) = p \circ \Phi(u_m, x_k) = f(x_k)$$

$$\Rightarrow |(\widehat{p} - p) \circ \Phi(u_m, x_k)| = |\widehat{p} \circ \Phi(\widehat{u}_m, x_k) - \widehat{p} \circ \Phi(u_m, x_k)|$$

$$\text{(Lipschitz assumption)} \qquad \leq C(B)\|\widehat{p}\|_{\mathcal{P}} |\Phi(\widehat{u}_m, x_k) - \Phi(u_m, x_k)|$$

Suppose $\Phi(u, x) = (x, u(x), \partial^{\boldsymbol{a}} u(x))$ for some multi-index $\boldsymbol{a}$ s.t. $|\boldsymbol{a}| \leq k \in \mathbb{N}$

$$|(\widehat{p} - p) \circ \Phi(u_m, x_k)| \lesssim C(B)\|\widehat{p}\|_{\mathcal{P}} \|\widehat{u}_m - u_m\|_{C^k(\Omega)}$$

$$\text{(Sobolev embedding)} \qquad \lesssim C(B)\|\widehat{p}\|_{\mathcal{P}} \|\widehat{u}_m - u_m\|_{H^{\gamma'}(\Omega)}$$

$$\text{(From filtering bound)} \qquad \lesssim C(B)\|\widehat{p}\|_{\mathcal{P}} \rho^{\gamma - \gamma'} (\|p\|_{\mathcal{P}} + \textstyle\sum_m \|u_m\|_{\mathcal{U}})$$

Controlling the Equation Error

$$|(\widehat{p} - p) \circ \Phi(u_m, x_k)| \lesssim C(B) \|\widehat{p}\|_{\mathcal{P}} \, \rho^{\gamma - \gamma'} \left(\|p\|_{\mathcal{P}} + \sum_m \|u_m\|_{\mathcal{U}} \right)$$

- Extra lemma: $\|\widehat{p}\|_{\mathcal{P}} \leq \|p\|_{\mathcal{P}} + \rho^\gamma \sum_m \|u_m\|_{\mathcal{U}^2}^2$ where $\|\cdot\|_{\mathcal{U}^2}$ is stronger norm on $\mathcal{U}$.
- Up to leading order

$$|(\widehat{p} - p) \circ \Phi(u_m, x_k)| \lesssim C(B) \|p\|_{\mathcal{P}} \, \rho^{\gamma - \gamma'} \left(\|p\|_{\mathcal{P}} + \sum_m \|u_m\|_{\mathcal{U}} \right)$$

- Apply sampling inequality for p and $\widehat{p}$ with fill distance

$$\varrho(B) := \sup_{s' \in B} \inf_{m,k} |s' - \Phi(u_m, x_k)|$$

$$\|\widehat{p} - p\|_{L^\infty(\Omega)} \leq C(B) \left(\varrho(B)^{\eta - Q/2} + \rho^{(\gamma - \gamma')} \right) \left(\|p\|_{\mathcal{P}} + \sum_m \|u_m\|_{\mathcal{U}} \right)$$

Thank you

Yasamin Jalalian, Juan Felipe Osorio Ramirez, Alex Hsu,, Bamdad Hosseini, and Houman Owhadi, **Data-Efficient Kernel Methods for Learning Differential Equations and Their Solution Operators: Algorithms and Error Analysis**, arXiv:2503.01036, 2025

References I

[1] S. L. Brunton, J. L. Proctor, and J. N. Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* 113.15 (2016), pp. 3932–3937. DOI: 10.1073/pnas.1517384113.

[2] N. B. Kovachki, S. Lanthaler, and A. M. Stuart. "Operator learning: Algorithms and analysis". In: *arXiv preprint arXiv:2402.15715* (2024).

[3] Z. Chen, Y. Liu, and H. Sun. "Physics-informed learning of governing equations from scarce data". In: *Nature communications* 12.1 (2021), p. 6136.

[4] F. Sun, Y. Liu, and H. Sun. "Physics-informed Spline Learning for Nonlinear Dynamics Discovery". In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Ed. by Z.-H. Zhou. Main Track. International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 2054–2061.

References II

[5] E Haber and D Oldenburg. “Joint inversion: a structural approach”. In: *Inverse problems* 13.1 (1997), p. 63.

[6] B. Kaltenbacher. “Regularization based on all-at-once formulations for inverse problems”. In: *arXiv [math.NA]* (Mar. 16, 2016).

[7] Y. Chen et al. “Solving and learning nonlinear PDEs with Gaussian processes”. In: *Journal of Computational Physics* 447 (2021), p. 110668.

[8] D. Long et al. “A kernel framework for learning differential equations and their solution operators”. In: *Physica D: Nonlinear Phenomena* 460 (2024), p. 134095.

[9] L. Zhang and H. Schaeffer. “On the convergence of the SINDy algorithm”. In: *Multiscale Modeling & Simulation* 17.3 (2019), pp. 948–972.

[10] P. Scholl et al. “The uniqueness problem of physical law learning”. In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2023, pp. 1–5.

References III

[11] Y. He, H. Zhao, and Y. Zhong. "How much can one learn a partial differential equation from its solution?" In: *Foundations of Computational Mathematics* 24.5 (2024), pp. 1595–1641.

[12] N. Boullé, D. Halikias, and A. Townsend. "Elliptic PDE learning is provably data-efficient". In: *Proceedings of the National Academy of Sciences* 120.39 (2023), e2303904120.