How much can one learn a PDE from its solution?

Hongkai Zhao
Duke University

Joint with Yuchen He and Yimin Zhong

Research partially supported by NSF DMS-2012860.

Learning

Learning is about modeling/understanding/approximating an "input-to-output" relation/function/mapping.

- Representation of the mapping is of crucial importance for understanding, accurate approximation, and computation efficiency.
- Learning from observation/measurement data has a long history.
- Understanding the underlying problem/model complexity is critical for effective learning.
  - how many degrees of freedom in the representation
  - how much data needed
  - the computation cost

PDE learning

- PDEs have a successful track record in modeling, studying, and predicting in science, engineering and many real world applications.
- Effectiveness comes from
  - a few terms that can capture various physical laws, diverse mechanisms, and rich dynamics
  - interpretable coefficients
  - insightful understanding
  - efficient computation
- Evolution PDE is an effective way to model dynamics or represent an ”initial-to-terminal” mapping.

PDE learning from its solution

Evolution PDE model

$$\partial_t u(x,t) = -\mathcal{L}u(x,t), \quad (x,t) \in \Omega \times [0,T], \;\; \Omega = \mathbb{T}^d$$
$$u(x,0) = u_0(x).$$

$$\mathcal{L}u(x) = \sum_{|\alpha|=0}^{n} p_\alpha(x)\partial^\alpha u(x),$$

Two approaches:

- Differential operator approximation (DOA) of the mapping $u(\cdot,t) \to u(\cdot,t+\Delta t)$ restricted to some finite dimensional space.
  - Pros: general and flexible.
  - Cons: large degrees of freedom, large amount of global solution data and expensive computation, difficult to decipher.
- Differential operator identification (DOI) built from differential operators and their functions in a given dictionary.
  - Pros: less degrees of freedom, less amount of and local data, less computation cost, return of an explicit PDE.
  - Cons: needs some prior knowledge for the dictionary.

Previous approaches and issues

A lot of approaches have been proposed, e.g., neural network based DOA, regression based DOI.

Issues:

- Global solution data on a dense space-time (from $t = 0$) grid are used.
- Use as many solution trajectories corresponding to as diverse initial data as one wants.

Previous approaches and issues

A lot of approaches have been proposed, e.g., neural network based DOA, regression based DOI.

Issues:

- Global solution data on a dense space-time (from $t = 0$) grid are used.
- Use as many solution trajectories corresponding to as diverse initial data as one wants.

In practice, solution data is available

- from one solution trajectory with uncontrollable initial data,
- observed by local sensors (for some time duration) at certain locations,
- after some time delay.

Basic questions for PDE learning

- How large is the data space spanned by all snapshots of a solution trajectory?
- Identifiability of PDEs using a single solution trajectory and stability.
- Data-driven and data-adaptive sampling.
- Robust PDE type identification using local measurements from a single trajectory.

Data space spanned by a solution trajectory: parabolic

Let $\mathcal{L}$ be a strongly elliptic operator of order $n = 2m$, then there is a $\mu > 0$ such that

$$e^{-\mathcal{L}_\mu t} = \frac{1}{2\pi i} \int_\Gamma e^{-zt}(z - \mathcal{L}_\mu)^{-1} dz \,, \quad \mathcal{L}_\mu = \mathcal{L} + \mu.$$

$\{\lambda_k, \phi_k\}_{k\geq 1}$ denote the eigenpairs of $\mathcal{L}_\mu$ sorted by $\Re\lambda_k$ in ascending order. $\{\lambda_k\}_{k\geq 1}$ satisfy the growth rate $\Re\lambda_k = O(k^\beta)$ with $\beta = n/d$.

Theorem 1

For $t \in [t_0 = \epsilon^r, T]$, there is an operator $\mathcal{L}_N = \sum_{k=-N}^{N} c_k e^{-z_k t}(z_k - \mathcal{L}_\mu)^{-1}$ with constants $c_k, z_k \in \mathbb{C}$ and $N = C(r, \mathcal{L}_\mu)|\log \epsilon|^2$

$$\|e^{-\mathcal{L}_\mu t} - \mathcal{L}_N(t)\|_{L_2(\Omega)\to L_2(\Omega)} \leq \epsilon.$$

Data space spanned by a solution trajectory: parabolic

Theorem 2

Assume $u_0(x) = \sum_{k=1}^{\infty} c_k \phi_k(x)$ *and* $|c_k| \leq \theta k^{-\gamma}$, $\theta > 0, \gamma > 1/2$, *then there exists a linear space* $V \subset L^2(\Omega)$ *of dimension* $C(\kappa, \mathcal{L}_\mu)|\log \epsilon|^2$

$$\|u(\cdot, t) - P_V u(\cdot, t)\| \leq C\epsilon \|u_0\|, \ \forall t \in [0, T].$$

P_V *is the projection operator onto* V, $C = C(\theta, \gamma)$, $\kappa = O(2\beta/(2\gamma - 1))$.

Data space spanned by a solution trajectory: parabolic

Theorem 2

Assume $u_0(x) = \sum_{k=1}^{\infty} c_k\phi_k(x)$ *and* $|c_k| \le \theta k^{-\gamma}$, $\theta > 0, \gamma > 1/2$, *then there exists a linear space* $V \subset L^2(\Omega)$ *of dimension* $C(\kappa, \mathcal{L}_\mu)|\log\epsilon|^2$

$$\|u(\cdot,t) - P_V u(\cdot,t)\| \le C\epsilon\|u_0\|, \ \forall t \in [0,T].$$

P_V *is the projection operator onto* V, $C = C(\theta,\gamma)$, $\kappa = O(2\beta/(2\gamma-1))$.

Key points: If $u_0(x) = \sum_{k=1}^{\infty} c_k\phi_k(x) \Rightarrow u(x,t) = e^{\mu t}\sum_{k=1}^{\infty} c_k e^{-\lambda_k t}\phi_k(x)$.

Take $M_\epsilon = O(\epsilon^{2/(1-2\gamma)})(\Rightarrow \|u(\cdot,t) - u_{M_\epsilon}(\cdot,t)\|_{L^2(\Omega)} \lesssim \epsilon)$, $L_\epsilon = O(|\log\epsilon|)$,

$$w_\epsilon(x,t) = \sum_{l=0}^{L_\epsilon} t^l \left(\sum_{k=1}^{M_\epsilon} c_k(-1)^l \frac{(\lambda_k-\mu)^l}{l!}\phi_k(x)\right) \in V_1 = \text{span}\left\{\sum_{k=1}^{M_\epsilon} c_k(-1)^l \frac{(\lambda_k-\mu)^l}{l!}\phi_k(x),\ 0 \le l \le L_\epsilon\right\}$$

For $t \in [0, \lambda_{M_\epsilon}^{-1}]$, $\quad \|u_{M_\epsilon}(x,t) - w_\epsilon(x,t)\| \lesssim \epsilon$.

For $t \in [\lambda_{M_\epsilon}^{-1} = O(\epsilon^{2\beta/(2\gamma-1)}), T]$,

$\exists V_2 = \text{span}\left\{(z_k - \mathcal{L}_\mu)^{-1}u_0, 1 = k \le N = O(|\log\epsilon|^2)\right\}$, $\|u(x,t) - P_{V_2}u(x,t)\| \lesssim \epsilon$.

$V = V_1 + V_2, \quad \dim V = O(|\log\epsilon|^2)$.

Data space spanned by a solution trajectory: hyperbolic

$$\partial_t u(x,t) + c(x)\cdot\nabla u(x,t) = 0, \quad (x,t)\in\Omega\times[0,T], \;\; \Omega = \mathbb{T}^d$$
$$u(x,0) = u_0(x).$$

Define the two correlation functions in space and time of a solution,

$$K(x,y) := \int_0^T u(x,s)u(y,s)ds, \quad (x,y)\in\Omega\times\Omega,$$
$$G(s,t) := \int_\Omega u(x,t)u(x,s)dx, \quad (s,t)\in[0,T]\times[0,T],$$

$K(x,y)$, $G(s,t)$ define two symmetric semi-positive compact integral operators on $L^2(\Omega)$ and $L^2[0,T]$ respectively, with the same non-negative eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_j \geq \ldots \to 0$.
Let V_K^k be the space spanned by the k leading eigenfunctions of $K(x,y)$,

$$\int_0^T \|u(\cdot,t) - P_{V_K^k}u(\cdot,t)\|^2_{L^2(\Omega)}dt = \min_{V\subset L^2(\Omega),\dim V=k}\int_0^T \|u(\cdot,t)-P_V u(\cdot,t)\|^2_{L^2(\Omega)}dt = \sum_{j=k+1}^{\infty}\lambda_j.$$

Data space spanned by a solution trajectory: hyperbolic

Theorem 3

If $c(x), u_0(x) \in C^p(\Omega)$, there exists a $V \subset L^2(\Omega)$ of dimension $O(\epsilon^{-2/p})$

$$\sqrt{\int_0^T \|P_V u(\cdot,t) - u(\cdot,t)\|^2_{L^2(\Omega)} dt} \le \epsilon.$$

Key point: $G(s,t) \in C^p([0,T]^2)$, its eigenvalue decays $\lambda_n = o(n^{-(p+1)})$.

Remark

- Explicit example can be constructed to show that if $u_0 \in C^p$, $\dim V \gtrsim \epsilon^{-1/(p+\frac{1}{2})}$.
- If both $c(x)$ and $u_0(x)$ are analytic, then the data space is of dimension $O(|\log \epsilon|^d)$.

Data space spanned by multiple solution trajectories

Theorem 4

If $\mathcal{L}$ is a self-adjoint strongly elliptic operator, then there exists a linear space $V \subset L^2(\Omega)$, $\dim V = O((\tau^{-1}|\log \epsilon|)^{d/n})$, for any solution $u(x,t)$ to the equation

$$\partial_t u = -\mathcal{L}u$$

with initial condition $u_0 \in L^2(\Omega)$,

$$\min_{f \in V} \|f(x) - u(x,t)\|_{L^2(\Omega)} \leq \epsilon \|u_0\|_{L^2(\Omega)}, \quad \forall t \in [\tau, T].$$

• For hyperbolic operator with multiple trajectories, the solution data space on an interval $[0, T]$ is as rich as the solution data space on $[\tau, T+\tau]$.

Potential challenges for PDE learning from its solution

DOA for the mapping: $u(\cdot, 0) \in H_p(\Omega) :\longrightarrow u(\cdot, \Delta t)$.

- With ϵ tolerance, $\dim(H_p(\Omega)) = O(\epsilon^{-d/p})$.
- Large number of parameters are needed in the approximation.
- Limited data in practice.

Differential operator identification (DOI)

Goal: identify a differential operator

- built from a dictionary of basic operators and their functions
- globally consistent
- using as few terms as possible
- fitting measurements well

using minimal local data from a single solution trajectory.

Identifiability from a single solution trajectory

PDE with constant coefficients: $\partial_t u(x,t) = \sum_{|\alpha|=0}^{n} p_\alpha \partial^\alpha u(x)$.
Let $\widehat{u}(\zeta, t)$ be the Fourier transform of the solution $u(x,t)$.

$$\widehat{u}(t,\zeta) = \widehat{u}(0,\zeta)\exp\left(-(2\pi)^{-d/2}\sum_{|\alpha|=0}^{n} p_\alpha (i\zeta)^\alpha t\right)$$

Theorem 5

Let $Q = \{\zeta \in \mathbb{Z}^d : \widehat{u}_0(\zeta) \neq 0\}$, if $|Q| \geq \max\left(\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor}\binom{2k+d-1}{d-1}, \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor}\binom{2k+d}{d-1}\right)$ and Q is not located on an algebraic polynomial hypersurface of degree $\leq n$ consists of only even order terms or odd order terms, then the PDE is uniquely determined by the solution at two instants $u(x,t_2), u(x,t_1)$ if $|t_2 - t_1|$ is small enough.

Identifiability from a single solution trajectory

PDE with variable coefficients: $\partial_t u(x,t) = \sum_{|\alpha|=0}^{n} p_\alpha(x)\partial^\alpha u(x)$.

- different regression problems at different locations,
- variation of the coefficients are intertwined with the solution in both frequency and spatial domain.

Lemma 6

Let $m = \binom{n+d}{d}$. For any given $x \in \Omega$, the parameters $p_\alpha(x)$ can be recovered if and only if one can find m instants $t_1, \ldots, t_m$ such that the matrix $A_{k,\alpha} := \partial^\alpha u(x, t_k)$ is non-singular.

Consider the limiting case that $t_1, \ldots, t_m \to 0$,
$A_{k,\alpha}$ becomes $S_{k,\alpha} := \partial^\alpha \mathcal{L}^{k-1} u_0(x)$, $k = 1, 2, \ldots, m$.

Theorem 7

Assume $u_0 = \sum_{j=1}^{r} w_j e^{i\zeta_j \cdot x}$, $r > \binom{(m-1)n+d}{d}$, and $\zeta_j \in \mathbb{Z}^d$ are not on an algebraic hypersurface of degree $(m-1)n$, $w_j \sim \mathcal{U}[a_j, b_j]$. If $\sum_{\alpha \in D_n} |p_\alpha(x)|^2 \neq 0$ almost everywhere in Ω, then the matrix S is non-singular almost surely wrt $\mathbb{P}$, the induced probability measure on $\Omega \times \prod_{j=1}^{m} [a_j, b_j]$.

Potential instability for DOI for parabolic PDE

Small perturbation: $p_\alpha(x) \to p_\alpha(x) + \mu f_\alpha(x)$, $|\mu| \ll 1$. $w(x,t) = \partial_\mu u(x,t)$ satisfies

$$\partial_t w(x,t) = \sum_{|\alpha|=0}^{n} p_\alpha(x)\partial^\alpha w(x,t) + \sum_{|\alpha|=0}^{n} f_\alpha(x)\partial^\alpha u(x,t),$$

$$w(x,0) = 0.$$

For parabolic PDE

$$\|w(x,t)\|^2_{L^2(\Omega\times[0,T])} \le c \sum_{0\le|\alpha|,|\beta|\le n} \int_\Omega K_{\alpha\beta}(x) f_\alpha(x) f_\beta(x)dx \le c\int_\Omega \lambda_m(x) \sum_{|\alpha|=0}^{n} |f_\alpha(x)|^2 dx$$

$K(x) = (K_{\alpha\beta}(x) = \int_0^T \partial^\alpha u(x,t)\partial^\beta u(x,t)dt)$ is a semi PD matrix for each $x \in \Omega$ and $\lambda_m(x) \ge 0$ is the smallest eigenvalue.

For short time $T \ll 1$, $\lambda_m(x) \le \frac{T^m}{m!}\|M(x)\| + O(T^{m+1})$

$$\|M(x)\| \le \left(\sum_{|\alpha|=0}^{n} |\partial^\alpha u_0(x)|^2\right)^{1/2} \left(\sum_{|\alpha|=0}^{n} |\partial^\alpha \mathcal{L}^{m-1} u_0(x)|^2\right)^{1/2}$$

$$\Rightarrow \min_{f_\alpha \in F} \|w\|^2_{L^2(\Omega\times[0,T])} \le \frac{cT^m}{m!} \|\mathcal{L}^{m-1}u_0\|_{W^{n,2}(\Omega)} \|u_0\|_{W^{n,2}(\Omega)} + O(T^{m+1}).$$

Potential instability for DOI for parabolic PDE

$\mathcal{L} = -\sum_{\alpha=0}^{n} p_\alpha \partial^\alpha$: an elliptic operator with constant coefficients in 1D.

$$u(x,t) = \sum_{k\in\mathbb{Z}} \phi_k(t) e^{ik\cdot x}, \quad \lambda_k = -\sum_{\alpha=0}^{n} p_\alpha (ik)^\alpha, \quad \phi_k(t) = c_k e^{-\lambda_k t}$$

$\widetilde{\mathcal{L}} = \mathcal{L} - \delta e^{iq\cdot x}\partial^{\alpha'}$, for some $q \in \mathbb{Z}$.

$$\tilde{u}(x,t) = \sum_{k\in\mathbb{Z}} \widetilde{\phi}_k(t) e^{ik\cdot x}, \quad \widetilde{\phi}_k(t) = e^{-\lambda_k t}\left(c_k + \int_0^t e^{\lambda_k s}\delta(i(k-q))^{\alpha'}\widetilde{\phi}_{k-q}(s)ds\right)$$

Theorem 8

Under the assumption $\Re\lambda_k \geq C_1\langle k\rangle^n$, $|c_k| \leq C_2\langle k\rangle^{-\beta}$, $\beta > \frac{1}{2}$, $\delta^2 T < \gamma C_1$, $\gamma \in (0,1)$

$$\|u(x,t) - \tilde{u}(x,t)\|^2_{L^2(\Omega\times[0,T])} \leq \frac{\gamma}{2(1-\gamma)}\left(\langle q\rangle^{-n}(2^n + \frac{(C_2)^2}{C_1})C_3 + \langle q\rangle^{2\alpha'-2\beta-n}2^{n+2\beta-2\alpha'}C_4\right).$$

$\langle k\rangle = (1+k^2)^{1/2}$, $C_3 = \sum_{k\in\mathbb{Z}}(1+k^2)^{\alpha'-\beta-n/2}$, $C_4 = \sum_{k\in\mathbb{Z}}(1+k^2)^{-n/2}$.

$\Rightarrow$ instability for high frequency perturbation.

Consistent and Sparse Local Regression(CaSLR) for DOI

Key features of CaSLR:

- enforcing global consist PDE
- using fewest possible terms
- fitting local patch data from a single solution trajectory

for identification of PDE with variable coefficients.

Remark. CaSLR has similarities to the method proposed by Rudy (2019) et al.

CaSLR

Assumption:

- PDE: $u_t(x,t) = \sum_{k=1}^{K} c_k(x,t) f_k(x,t)$, f_k's are terms from a dictionary.
- Solution data on local patches Ω_j centered at (x_j, t_j), $j = 1, \ldots, J$.
- Patch size can resolve the variation of $c_k(x,t)$.

- Define the local regression error in each patch Ω_j.

$$\mathcal{E}_j^{loc}(\hat{\mathbf{c}}_j) = \sum_{(x_{j,m}, t_{j,m}) \in \Omega_j} \left(u_t(x_{j,m}, t_{j,m}) - \sum_{k=1}^{K} \hat{c}_k^j f_k(x_{j,m}, t_{j,m}) \right)^2 .$$

- Define the global regression error

$$\mathcal{E}(\hat{\mathbf{c}}) = \sum_{j=1}^{J} \mathcal{E}_j^{loc}(\hat{\mathbf{c}}_j), \quad \hat{\mathbf{c}} = [\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_J], \quad \hat{\mathbf{c}}_j = [\hat{c}_1^j, \ldots, \hat{c}_K^j].$$

- For each l, find $\hat{\mathbf{c}}^l = \arg\min_{\hat{\mathbf{c}}} \mathcal{E}(\hat{\mathbf{c}})$ subject to: $||\hat{\mathbf{c}}||_{\text{Group}-\ell_0} = l$
  $||\hat{\mathbf{c}}||_{\text{Group}-\ell_0} = ||(||\hat{c}_1||_1, \ldots, ||\hat{c}_K||_1)||_0$, using Group Subspace Pursuit (G-SP) proposed by He (2022) et al.

- Find l minimizes $S^l = \mathcal{E}(\hat{\mathbf{c}}^l) + \rho \frac{l}{K}$.

Data-driven and data-adaptive sampling

Not all data are equal. ⇒ Not all patches should be used indiscretely.

- CaSLR is based on consistent PDE identification on local patches.
- Ideal sensor: small patch size with fine resolution that detects diverse modes.
- Bad patch data: 1) patch data with little mode content cause instability. 2) patch data with rapid change induce large errors.

Numerical local Sobolev semi-norm can be used to filter out bad patches.

$$\beta_{ij} = \sqrt{\frac{1}{m_j}\sum_{m=1}^{m_j}\sum_{p=1}^{P_{\max}}(\partial_x^p u(x_{j,m}, t_{j,m}))^2}$$

Remark.

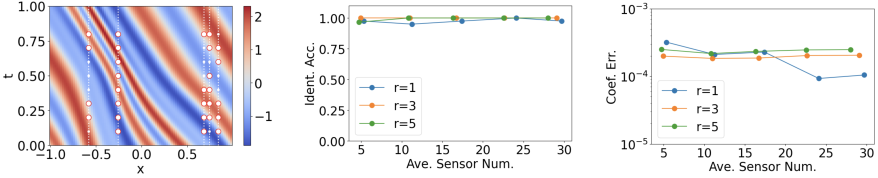
- When data contains noise more sophisticated process is needed.
- Once PDE type is determined, more accurate reconstruction of coefficients can be achieved.

Experiments

Example 1: Transport equation.

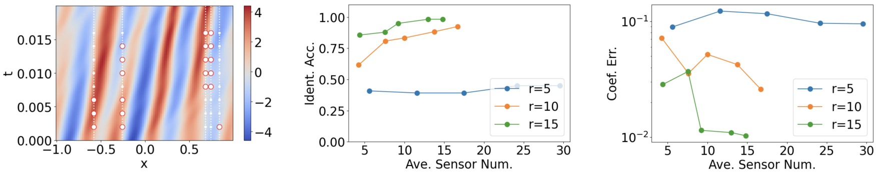
$$u_t(x,t) = (1 + 0.5\sin(\pi x)\tau(t; -10, 0.5))u_x(x,t)\ , \quad (x,t) \in [-1,1) \times (0,1],$$
$$u(x,0) = \sin(4\pi(x+0.1)) + \sin(6\pi x) + \cos(2\pi(x-0.5)) + \sin(2\pi(x+0.1)).$$

where $\tau(t; s, t_c) = 0.5 + 0.5\tanh(s(t - t_c))), \quad t \in \mathbb{R}.$



Example 2: KdV type equation.

$$u_t(x,t) = (3 + 200t\sin(\pi x))u(x,t)u_x(x,t) + \frac{5+\sin(\frac{400\pi t}{3})}{100}u_{xxx}(x,t),$$
$$u(x,0) = \sin(4\pi(x+0.1)) + 2\sin(5\pi x) + \cos(2\pi(x-0.5)) + \sin(3\pi x) + \cos(6\pi x).$$

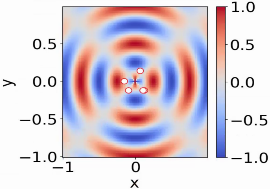


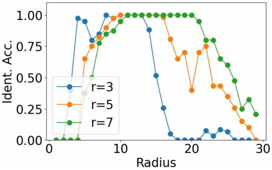
Experiments

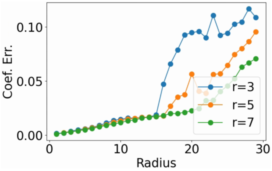
Example 3: 2D circular flow.

$$u_t(x,y,t) = -yu_x(x,y,t) + xu_y(x,y,t)$$
$$u(x,y,0) = f(x,y)$$

$u(x,y,t) = f(\sqrt{x^2+y^2}\cos(\arctan(y/x)-t), \sqrt{x^2+y^2}\sin(\arctan(y/x)-t))$,
where $f(x,y) = \cos(4\sqrt{x^2+y^2})\cos(2\arctan(y/x))$.



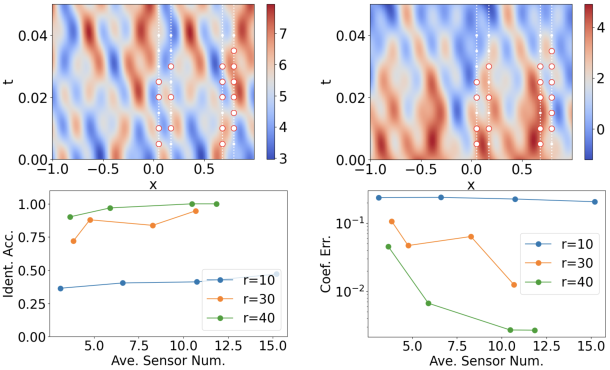




Experiments

Example 4: Schrödinger equation.

$$i\psi_t = \frac{1}{2}\psi_{xx} - V(x,t)\psi, \quad V(x,t) = -10 - 2\sin(40\pi t)\cos(\pi x).$$



Thank you for your attention!