How much can one learn a PDE from its solution?

Presented by Kyra Cho. Initially presented by Yimin Zhong from Auburn University, during the applied math colloquium on 4/9/24.

Objectives

- **Show how the approximate dimension of the data space by all the snapshots along a solution trajectory depends on a PDE's differential operator and initial data.**
- **Study the identifiability of a differential operator from solution data on local patches.**
- **Propose a consistent and sparse local regression method for general PDE identification. The method should correctly identify the PDE using a single trajectory and using a minimal amount of data.**

The original presentation focuses on those three objectives. However, only the implementation and analysis of the main method will be discussed in this short presentation.

Introduction

PDEs are an effective way to model phenomena in science, engineering, and many real-world applications. Its effective comes from the fact that it can capture various physical laws, mechanisms, and dynamics. The main question of this study is whether one can learn a PDE model directly from measured data, its derivatives, integrals, and other transformed quantities. And the particular method studied here is the differential operator identification (DOI) method, which determines the explicit form of differential operators by choosing terms from a predetermined dictionary to approximate the underlying PDE. The CaSLR method is used with data which is the solution corresponding to one initial condition and existing over some time durations.

The CaSLR method

Suppose the solution data is available at the neighborhoods of different locations, i.e., local patches, that can be used to approximate the solution derivatives and their functions corresponding to those terms in the dictionary. Using the measured/computed solution and its derivatives at certain locations, one can construct the CaSLR method. Assume the unknown PDE takes the following form:

$$u_t(x,t) = \sum_{k=1}^{K} c_k(x,t) f_k(x,t),$$

where

$$\mathcal{F} = \{f_k : \Omega \times [0,T] \to \mathbb{R}\}_{k=1}^{K}$$

is a set of functions (a dictionary of features) that contains partial derivatives of u with respect to x, functions of u such as sin(u) and u squared, and

$$c_k : \Omega \times [0,T] \to \mathbb{R}, \quad k = 1,2,\dots,K$$

are the respective coefficients. One can approximate the PDE using constants in each small neighborhood centered at $(x_j,t_j), j = 1,2,\dots,J$.

Denote Ω_j to be the local neighborhood centered at

$$(x_j,t_j) \text{ and } \hat{\mathbf{c}}_j = (\hat{c}_1^j,\dots,\hat{c}_K^j),$$

And define the local regression error in each patch as

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

The global regression error is defined as

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

For each $l = 1,2,\dots,K,$ one searches for terms in the dictionary whose linear combination using $\hat{\mathbf{c}}_j$ minimizes the global fitting error.

Specifically, one optimizes

$$\hat{\mathbf{c}}^l = \arg\min_{\hat{\mathbf{c}}} \mathcal{E}(\hat{\mathbf{c}})$$
$$\text{subject to:} \quad \|\hat{\mathbf{c}}\|_{\text{Group}-\ell_0} = l$$

using the Group Subspace Pursuit method. The GSP method identifies groups of functions from the dictionary that best fit the residual obtained from the previous iteration. And GSP exploits the structured sparsity to recover sparse data with improved accuracy compared to Subspace Pursuit, making it useful in this situation because data is available in patches. And furthermore, the group sparsity is defined by

$$\|\hat{\mathbf{c}}\|_{\text{Group}-\ell_0} = \|(\|\tilde{\mathbf{c}}_1\|_1,\dots,\|\tilde{\mathbf{c}}_K\|_1)\|_0,$$

where $\tilde{\mathbf{c}}_k = (\hat{c}_k^1,\dots,\hat{c}_k^J) \in \mathbb{R}^J, k = 1,2,\dots,K.$

As the sparsity level increases, the global regression error decreases. And furthermore, the model score is defined as

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

where $\rho > 0$ is a penalty parameter for using more terms form the dictionary.

The candidate with l^* features is the optimal if

$$S^{l^*} = \min_{l=1,\dots,K-1} S^l.$$

The model score is a metric that evaluates a candidate model by considering two factors: the fitting error and the model complexity penalty rho. Here, we fix it to be the mean of $\{\mathcal{E}(\hat{\mathbf{c}}^k)\}_{k=1}^{K}$ so that the two components of the model score are balanced.

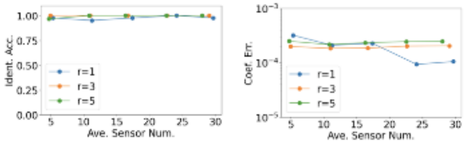
Numerical Implementation

In this section, the CaSLR method is implemented on well-known PDEs and its accuracy is compared using the Jacquard score.

$$J(S_0,S_1) = \frac{|S_0 \cap S_1|}{|S_0 \cup S_1|}$$

Example 1: Transport equation.

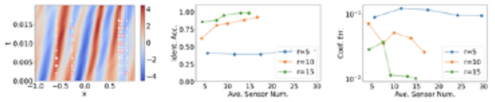
$$u_t(x,t) = (1 + 0.5\sin(\pi x)\tau(t;-10,0.5))u_x(x,t), \quad (x,t) \in [-1,1) \times (0,1],$$
$$u(x,0) = \sin(4\pi(x+0.1)) + \sin(6\pi x) + \cos(2\pi(x-0.5)) + \sin(2\pi(x+0.1)).$$



The method on the transport equation is very reliable even with a single sensor whose patch size is small. This is due to the solution data containing diverse Fourier modes over the space-time domain. According to He et al, a signal with many Fourier modes has redundant information across different frequency components, and the redundancy acts as a form of regularization, making GSP more stable and less susceptible to noise or errors.

Example 2: KdV type equation.

$$u_t(x,t) = (3 + 200t\sin(\pi x))u(x,t)u_x(x,t) + \frac{5+\sin(\frac{400\pi t}{3})}{100}u_{xxx}(x,t), \quad (x,t) \in [-1,1) \times (0,1.5 \times 10^{-2}],$$
$$u(x,0) = \sin(4\pi(x+0.1)) + 2\sin(5\pi x) + \cos(2\pi(x-0.5)) + \sin(3\pi x) + \cos(6\pi x).$$



And compared to the transport equation, patches of larger size are needed for accurate identification. This is because there are more terms in the PDE and of different orders, so more low-frequency Fourier modes for the lower order terms in the PDE need to be included in each patch.

Example 3: Schrödinger equation.

$$i\psi_t = \frac{1}{2}\psi_{xx} - V\psi$$

Using the potential function

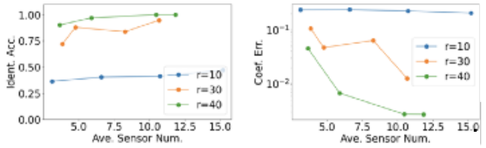
$$V = -10 - 2\sin(40\pi t)\cos(\pi x)$$

and letting $\psi = u + iv$, we have

$$\begin{cases} u_t(x,t) &= \frac{1}{2}v_{xx}(x,t) - V(x,t)v(x,t), \\ v_t(x,t) &= -\frac{1}{2}u_{xx}(x,t) + V(x,t)u(x,t). \end{cases}$$

We let the initial conditions be $u(x,0) = 5 + f_1(x)$

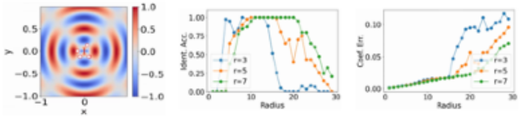
and $v(x,0) = 3 + f_2(x)$,i.e., random functions with four modes, plus a constant.



And the graphs show that identification accuracy and coefficient of error improves proportionally with the average sensor number and the radii of the patches.

Example 4: 2-D circular flow.

$$u_t(x,y,t) = -yu_x(x,y,t) + xu_y(x,y,t), \quad (x,y,t) \in \mathbb{R}^2 \times (0,2\pi]$$
$$u(x,y,0) = f(x,y), \quad (x,y) \in \mathbb{R}^2$$

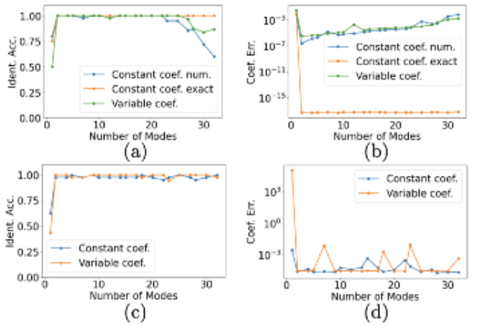


Here, a deterministic initial condition was used. 'r' is the sensing radius of the sensors, and 'Radius' is the radius of the circle in which the sensors are randomly located. When the sensors are close to the origin, the PDE type is hard to identify because the coefficients are all close to zero and the solution is almost constant 1.

The low reconstruction error is misleading because it means that recovered coefficients, including those for wrong features, are close to zeros. On the other hand, when the sensors are located far away from the origin the low frequency Fourier modes cannot be detected when the local patch is small. The PDE identification and the coefficient error are improved when the patch size becomes larger. The best result is seen when the sensors are at a moderate distance from the origin.

Exploration: Identification with random initial conditions.

I. transport equation with constant speed: $u_t(x,t) = 2u_x(x,t)$;
II. transport equation with variable speed: $u_t = (2+\sin(2\pi t))u_x(x,t)$;
III. heat equation with constant coefficient: $u_t(x,t) = 0.5u_{xx}(x,t)$;
IV. heat equation with variable coefficient: $u_t(x,t) = (0.5+0.25\sin(2\pi t))u_{xx}(x,t)$.



Graphs (a) and (b) correspond to the transport equations, and graphs (c) and (d) correspond to the heat equations. In all cases, the identification accuracy becomes correct when the initial condition has at least 2 modes, since there's only one perm in the PDE. For the transport equations, significant numerical errors emerge when the initial data contain high frequency modes that the grid can't resolve. On the other hand, this isn't true for diffusion equations since the solution is smoothed out in time quickly. Also, the coefficient errors for the variable coefficients are greater than with constant coefficients. This is because the PDE is approximated in each local patch by a PDE with constant coefficients.

Conclusion

These experiments show that the diversity of Fourier modes in the solution data are important for accuracy. On the other hand, if the solution varies too quickly, then identification is compromised due to numerical errors. The CaSLR method is globally consistent and involves as few terms as possible, using a single solution trajectory.

References

He, Y., Zhao, H. & Zhong, Y. How Much Can One Learn a Partial Differential Equation from Its Solution?. https://doi.org/10.1007/s10208-023-09620-z

Y. He, S.-H. Kang, W. Liao, H. Liu, and Y. Liu. Group projected subspace pursuit for identification of variable coefficient differential equations (GP-IDENT). arXiv preprintarXiv:2304.05543, 2023.