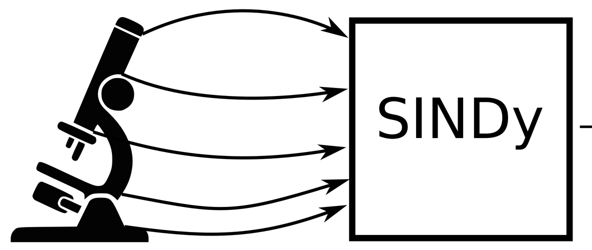
SINDy for model discovery in active systems

Christopher Amey[1], Chaitanya Joshi[2], Aparna Baskaran[1], Michael Hagan[1]

[1] Department of Physics, Brandeis University, [2] Department of Physics, Tufts University

What is SINDy?

Sparse identification of non-linear dynamics (SINDy) is an advanced regression technique that takes raw experimental or numerical data as input and returns a sparse best-fit PDE model.



$$\partial_t Q = -0.94u.\nabla Q - 0.96(\Omega.Q - Q.\Omega) + 0.95E - 2.1(E:Q)Q - 0.68\mathrm{Tr}[Q^2]Q$$

wikimedia commons

How it works

Consider a fluid system where we have velocity and density data, $\boldsymbol{u}(\boldsymbol{x})$ and $\rho(\boldsymbol{x})$, at space-time coordinates $\boldsymbol{x} = \{x, y, z, t\}$, and we want a PDE that describes $\partial_t \boldsymbol{u}$. Without further information, the underlying PDE will take the form

$$\partial_t \boldsymbol{u} = \sum_{n=0}^{N} \xi_n f_n(\boldsymbol{u}, \boldsymbol{\rho}, \nabla \boldsymbol{u}, \nabla \rho, \ldots),$$

where f_n are different terms with coefficients ξ_n.

Steps for SINDy:

1. Create a library of relevant PDE terms, f_n (requires some knowledge of system)
2. Numerically differentiate and sample data
3. Perform *sparse* regression to find ξ_n $\Rightarrow$ most coefficients will be zero!

For example, a simple library could be of the form,

$$\partial_t \boldsymbol{u} = \xi_0 \boldsymbol{u} + \xi_1 \nabla \rho + \xi_2 \rho \boldsymbol{u} + \ldots + \xi_N \boldsymbol{u}^9$$

If we sample M different space-time points,

$$\begin{bmatrix} \partial_t \boldsymbol{u}(\boldsymbol{x}_0) \\ \partial_t \boldsymbol{u}(\boldsymbol{x}_1) \\ \vdots \\ \partial_t \boldsymbol{u}(\boldsymbol{x}_M) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}(\boldsymbol{x}_0) & \nabla \rho(\boldsymbol{x}_0) & \cdots & \boldsymbol{u}^9(\boldsymbol{x}_0) \\ \boldsymbol{u}(\boldsymbol{x}_1) & \nabla \rho(\boldsymbol{x}_1) & \cdots & \boldsymbol{u}^9(\boldsymbol{x}_1) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}(\boldsymbol{x}_M) & \nabla \rho(\boldsymbol{x}_M) & \cdots & \boldsymbol{u}^9(\boldsymbol{x}_M) \end{bmatrix} \begin{bmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_N \end{bmatrix},$$

which, in simplified notation, is

$$U_t = \Theta \Xi,$$

where U_t is a length M vector, Θ is an $N \times M$ matrix, and Ξ is a length N vector of unknown coefficients. By performing sparse regression on Θ (e.g. ridge or lasso regression), we can obtain Ξ with most entries equal to zero, whereas standard least squares would overfit, resulting in non-zero coefficients even for terms that don't aren't important to the system.

2D active nematic simulations

We obtain data by numerically solving nametic hydrodynamic equations,

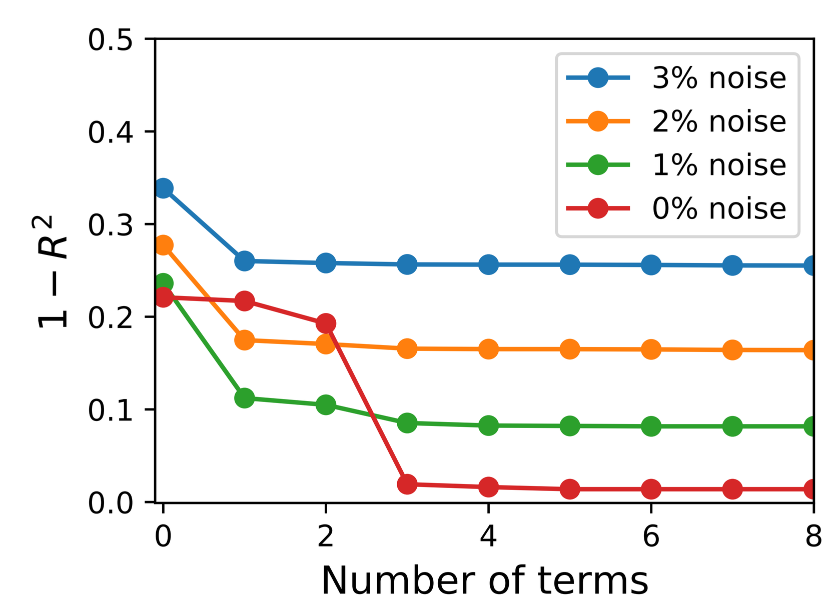
$$\partial_t \boldsymbol{Q} + \boldsymbol{u}.\nabla \boldsymbol{Q} = \boldsymbol{S}(\nabla \boldsymbol{u}, \boldsymbol{Q}) - D_R H$$
$$0 = -\gamma \boldsymbol{u} - \nabla p + \eta \nabla^2 \boldsymbol{u} - \alpha \nabla.\boldsymbol{Q}$$
$$\nabla.\boldsymbol{u} = 0,$$

where

$$H = \kappa \nabla^2 \boldsymbol{Q} - \left(a_1 \boldsymbol{Q} - a_2 \left(\boldsymbol{Q}^2 - \frac{1}{3} tr(\boldsymbol{Q}^2) \boldsymbol{I} \right) + a_3 tr(\boldsymbol{Q}^2) \boldsymbol{Q} \right)$$

$$\boldsymbol{S}(\nabla \boldsymbol{u}, \boldsymbol{Q}) = \boldsymbol{\Omega}.\boldsymbol{Q} - \boldsymbol{Q}.\boldsymbol{\Omega} + \lambda(\boldsymbol{Q}.\boldsymbol{E} + \boldsymbol{E}.\boldsymbol{Q}) + \frac{2\lambda}{3} \boldsymbol{E} - \frac{2}{3} \lambda (\boldsymbol{E} : \boldsymbol{Q}) \boldsymbol{I}$$

The reliability of SINDy can be tested by adding artificial noise to the data and analyzing how the accuracy of the fit changes. The noise is multiplicative and is drawn from a normal distribution with mean 1 and standard deviation σ, where 1% noise corresponds to $\sigma = 0.01$.

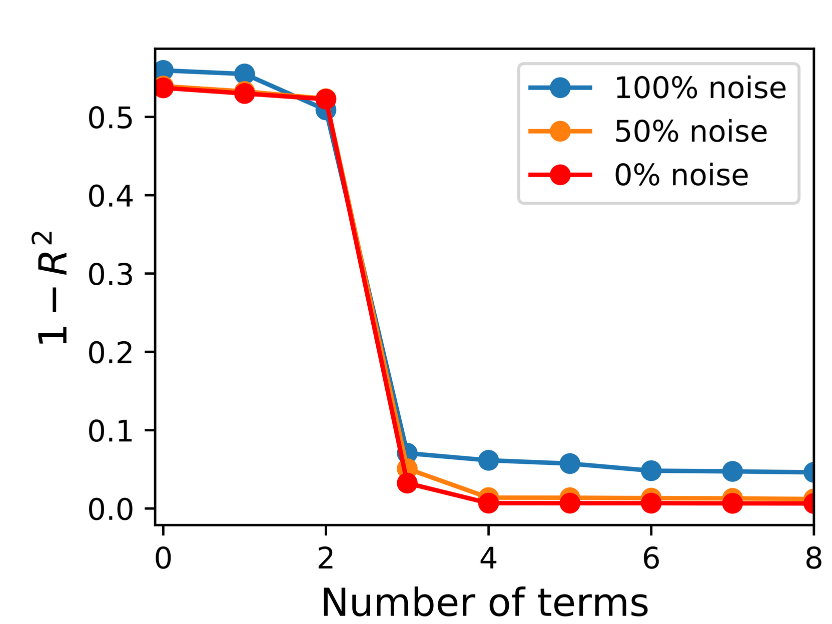


- With clean data, SINDy works!
- Modest noise (3%) renders SINDy ineffective
- Even with only 1% noise, high order terms are lost

Noise mitigation

Averaging: We can reduce noise in our measurements by integrating over a small window, Ω, with a distribution or "test function", $\omega(\boldsymbol{x})$. Returning to our example library, this would look like

$$\int_\Omega \omega \, \partial_t \boldsymbol{u} d\boldsymbol{x} = \int_\Omega \omega \left(\xi_0 \boldsymbol{u} + \xi_1 \nabla \rho + \xi_2 \rho \boldsymbol{u} + \ldots + \xi_N \boldsymbol{u}^9 \right) d\boldsymbol{x}$$



- Even 50% noise results in a reasonably good fit
- Easy to implement
- Suited for long-wavelength physics (hydrodynamics)

Weak form

If we choose our test function so that it has zero weight on the boundary of the integral window, $\partial\Omega$, then we can easily transfer derivatives to ω and off of noisy data via integration by parts. For example,

$$\int_\Omega \omega \, \partial_t \boldsymbol{u} d\boldsymbol{x} = \omega \boldsymbol{u}\big|_{\partial\Omega} - \int_\Omega \partial_t \omega \, \boldsymbol{u} d\boldsymbol{x} = -\int_\Omega \partial_t \omega \, \boldsymbol{u} d\boldsymbol{x},$$

where the second term is equal to zero since $\omega(\partial\Omega) = 0$.

In addition to reducing noise, using the weak form of a PDE gives lattitude to remove troublesome variables from equations. For instance, the pressure is a term in the Navier-Stokes equation but it is often impossible to measure experimentally. By choosing a divergence-free vector as our test function,

$$\int_\Omega \boldsymbol{\omega}.(\nabla p) \, d\boldsymbol{x} = \boldsymbol{\omega} p\big|_{\partial\Omega} - \int_\Omega p \, (\nabla.\boldsymbol{w}) \, d\boldsymbol{x} = 0,$$

which makes removal of certain immeasurable variables, such as pressure, possible.

Outlook

We have a useful technique that has been extensively tested on 2D simulation and experimental data by Joshi, et al. Moving forward, we will further develop SINDy for generic 3D active systems, including:

- Experimental active nematic liquid crystals
- Experimental active nematic elastomers
- Emergent physics from particle-based simulations

References

1) Joshi et al, Data-driven discovery of active nematic hydrodynamics, in preparation
2) Brunton et al, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, PNAS 2016
3) Brunton et al, Sparse Identification of Nonlinear Dynamics With Control, IFAC 2016
4) Alves and Fiuza, Data-driven discovery of reduced plasma physics models from fully-kinetic simulations, arXiv 2020
5) Reinbold et al, Using noisy or incomplete data to discover models of spatiotemporal dynamics, PRE 2020

Acknowledgements

This work is supported by: the Brandeis Center for Bioinspired Soft Materials

NSF MRSEC, DMR-2011846