SINDy for model discovery in active systems
Christopher Amey\textsuperscript{1}, Chaitanya Joshi\textsuperscript{2}, Aparna Baskaran\textsuperscript{1}, Michael Hagan\textsuperscript{1}
\textsuperscript{1} Department of Physics, Brandeis University, \textsuperscript{2} Department of Physics, Tufts University

\begin{center}
\textbf{What is SINDy?}
\end{center}

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.

\begin{center}
\textbf{How it works}
\end{center}

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

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

where $f_n$ are different terms with coefficients $\xi_n$.

\begin{center}
\textbf{Steps for SINDy:}
\end{center}

- Create a library of relevant PDE terms, $f_n$ (requires some knowledge of system)
- Numerically differentiate and sample data
- Perform sparse regression to find $\xi_n$

$\Rightarrow$ most coefficients will be zero!

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

$$\partial_t u = \xi_0 u + \xi_1 \nabla \rho + \xi_2 \rho u + \ldots + \xi_N u^N.$$  

If we sample $M$ different space-time points,

$$\begin{bmatrix}
\partial_t u(x_0) \\
\partial_t u(x_1) \\
\vdots \\
\partial_t u(x_M)
\end{bmatrix} = \begin{bmatrix}
u(x_0) \\
\nabla \rho(x_0) \\
\vdots \\
\nabla \rho(x_M)
\end{bmatrix}
\begin{bmatrix}
\xi_0 \\
\xi_1 \\
\vdots \\
\xi_N
\end{bmatrix},$$

which, in simplified notation, is

$$U_t = \Theta \Xi,$$

where $U_t$ is a length $M$ vector, $\Theta$ is an $N \times M$ matrix, and $\Xi$ is a length $N$ vector of unknown coefficients. By performing sparse regression on $\Theta$ (e.g., ridge or lasso regression), we can obtain $\Xi$ 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.

\begin{center}
\textbf{2D active nematic simulations}
\end{center}

We obtain data by numerically solving nematic hydrodynamic equations,

$$\begin{align*}
\partial_t Q + u \cdot \nabla Q &= S(\nabla u, Q) - D\partial_t H \\
\nabla \cdot u &= 0,
\end{align*}$$

where

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

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 $\sigma$, where 1% noise corresponds to $\sigma = 0.01$.

\begin{center}
\textbf{Noise mitigation}
\end{center}

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

$$\int_{\Omega} \omega \partial_t u dx = \int_{\Omega} \omega (\xi_0 u + \xi_1 \nabla \rho + \xi_2 \rho u + \ldots + \xi_N u^N) dx.$$

\begin{center}
\textbf{Weak form}
\end{center}

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 $\omega$ and off of noisy data via integration by parts. For example,

$$\int_{\Omega} \omega \partial_t u dx = \omega u|_{\partial \Omega} - \int_{\Omega} \partial_t \omega u dx = -\int_{\Omega} \partial_t \omega u dx,$$

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 latitude 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,

$$\int_{\Omega} \omega \cdot (\nabla p) dx = \omega p|_{\partial \Omega} - \int_{\Omega} \partial_t (\omega \cdot u) dx = 0,$$

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

\begin{center}
\textbf{Outlook}
\end{center}

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

\begin{center}
\textbf{References}
\end{center}

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, preprint 2016
4) Alves and Fiuza, Data-driven discovery of reduced plasma physics models from fully-lattice simulations, arXiv 2020
5) Reinhold et al. Using noisy or incomplete data to discover models of spatiotemporal dynamics, PRE 2020

\begin{center}
\textbf{Acknowledgements}
\end{center}

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

NSF MRSEC, DMR-2011846