Data requirements & Library

Filton workshop 2024

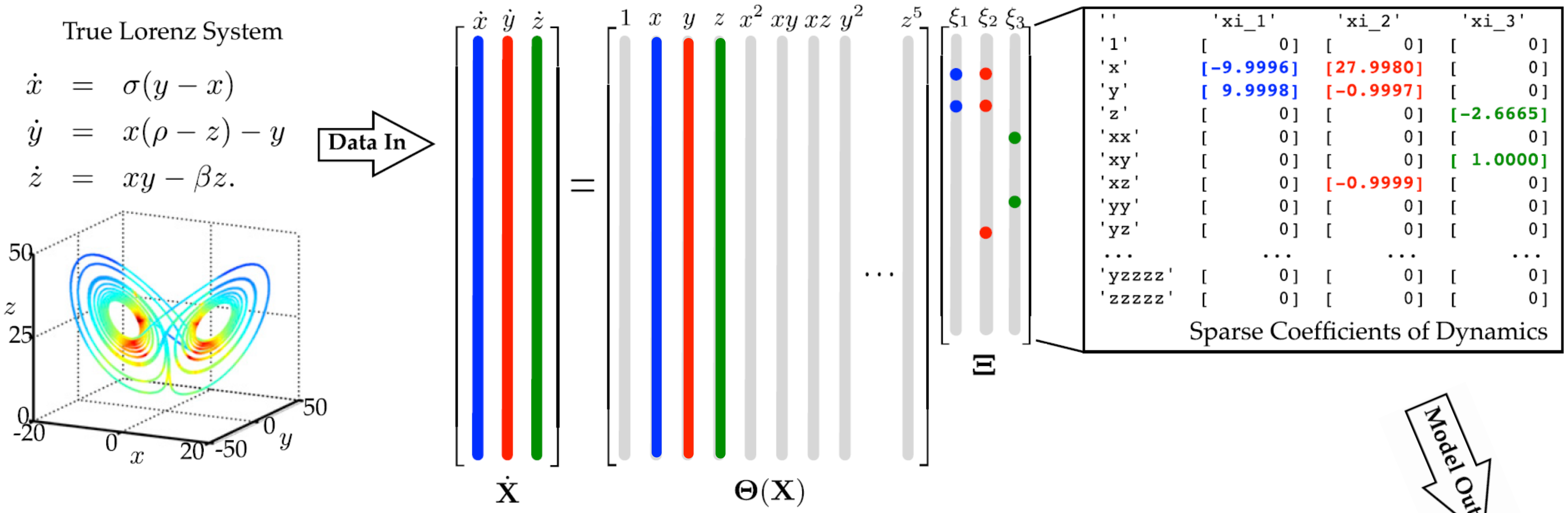
Urban Fasel

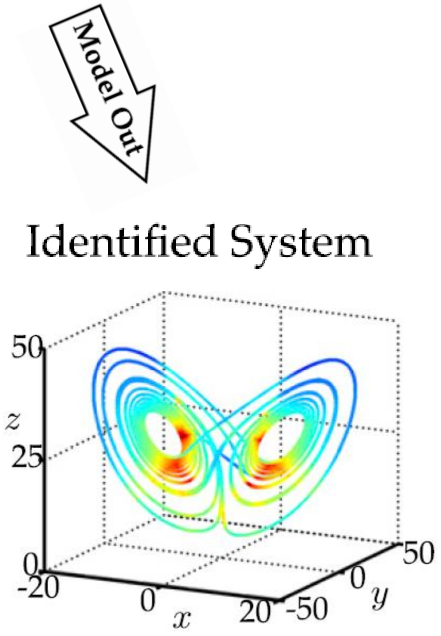
Imperial College London

Literature

- **Data requirements: sampling duration & rate:**
  - KP Champion, SL Brunton, JN Kutz (2019) Discovery of Nonlinear Multiscale Systems: Sampling Strategies and Embeddings.
- **Rational functions:**
  - NM Mangan, SL Brunton, JL Proctor, JN Kutz (2016) Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics.
  - K Kaheman, JN Kutz, SL Brunton (2020) SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics.
- **Curse of dimensionality**
  - K Champion, B Lusch, JN Kutz, SL Brunton (2019) Data-driven discovery of coordinates and governing equations.
  - P Gelß, S Klus, J Eisert, C Schütte (2019) Multidimensional Approximation of Nonlinear Dynamical Systems.
  - JC Loiseau, SL Brunton (2018) Constrained sparse Galerkin regression.
  - Y Guan, SL Brunton, I Novosselov (2021) Sparse nonlinear models of chaotic electroconvection.
  - A Kaptanoglu et al (2021) Promoting global stability in data-driven models of quadratic nonlinear dynamics.

SINDy – sparse regression





SINDy sparse regression

$$\rightarrow \quad \hat{\boldsymbol{\xi}}_k = \operatorname{argmin}_{\boldsymbol{\xi}_k} \underbrace{\left\|\dot{\mathbf{X}}_k - \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\xi}_k\right\|_2^2}_{\text{least squares}} + \underbrace{\lambda\|\boldsymbol{\xi}_k\|_0}_{\ell_0\text{-penalized}}$$

Tutorial outline

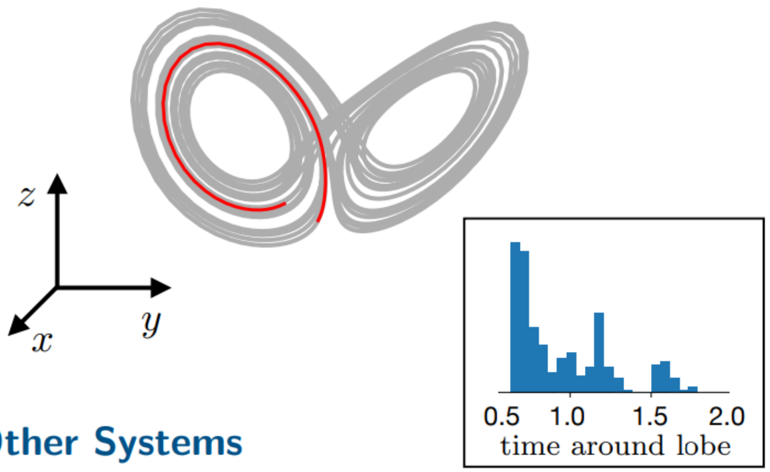
- **Data requirements**
  - Sampling duration & rate
  - Noise
  - Disambiguating multiple consistent models
- **Library**
  - Rational functions
  - Curse of dimensionality

Data requirements

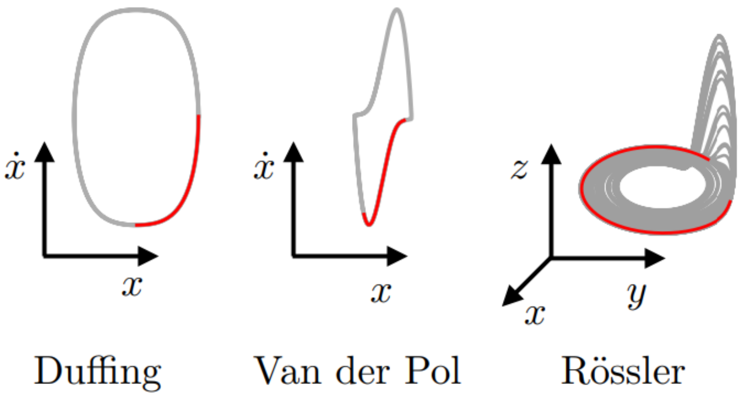
- Sampling duration & rate
- Noise
- Disambiguating multiple consistent models

Data requirements – Sampling duration & rate

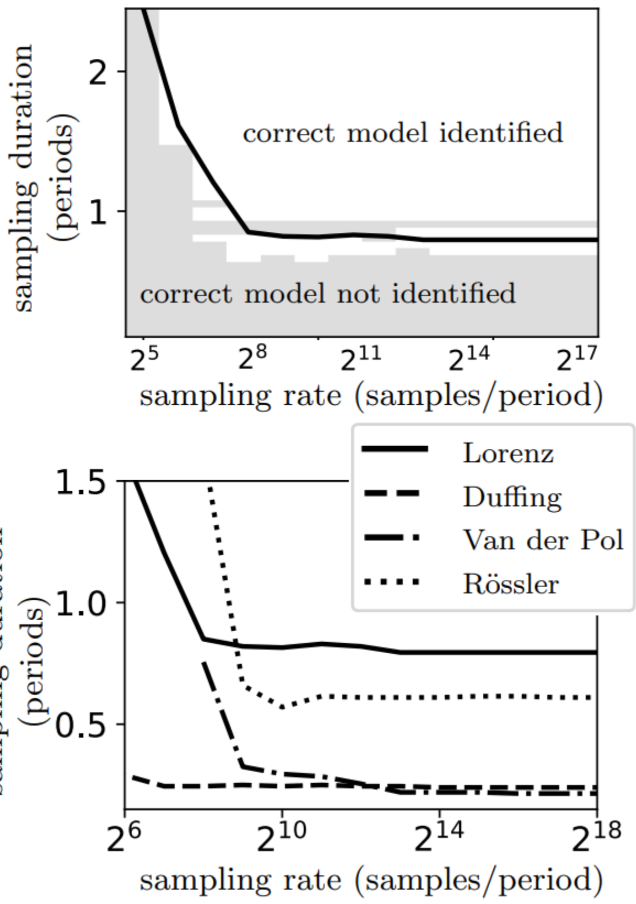
Lorenz System



Other Systems



Sampling Requirements



– attractor

– average portion of attractor sampled from to discover correct system (correct sparsity pattern)

KP Champion, SL Brunton, JN Kutz (2019) Discovery of Nonlinear Multiscale Systems: Sampling Strategies and Embeddings.

Duration

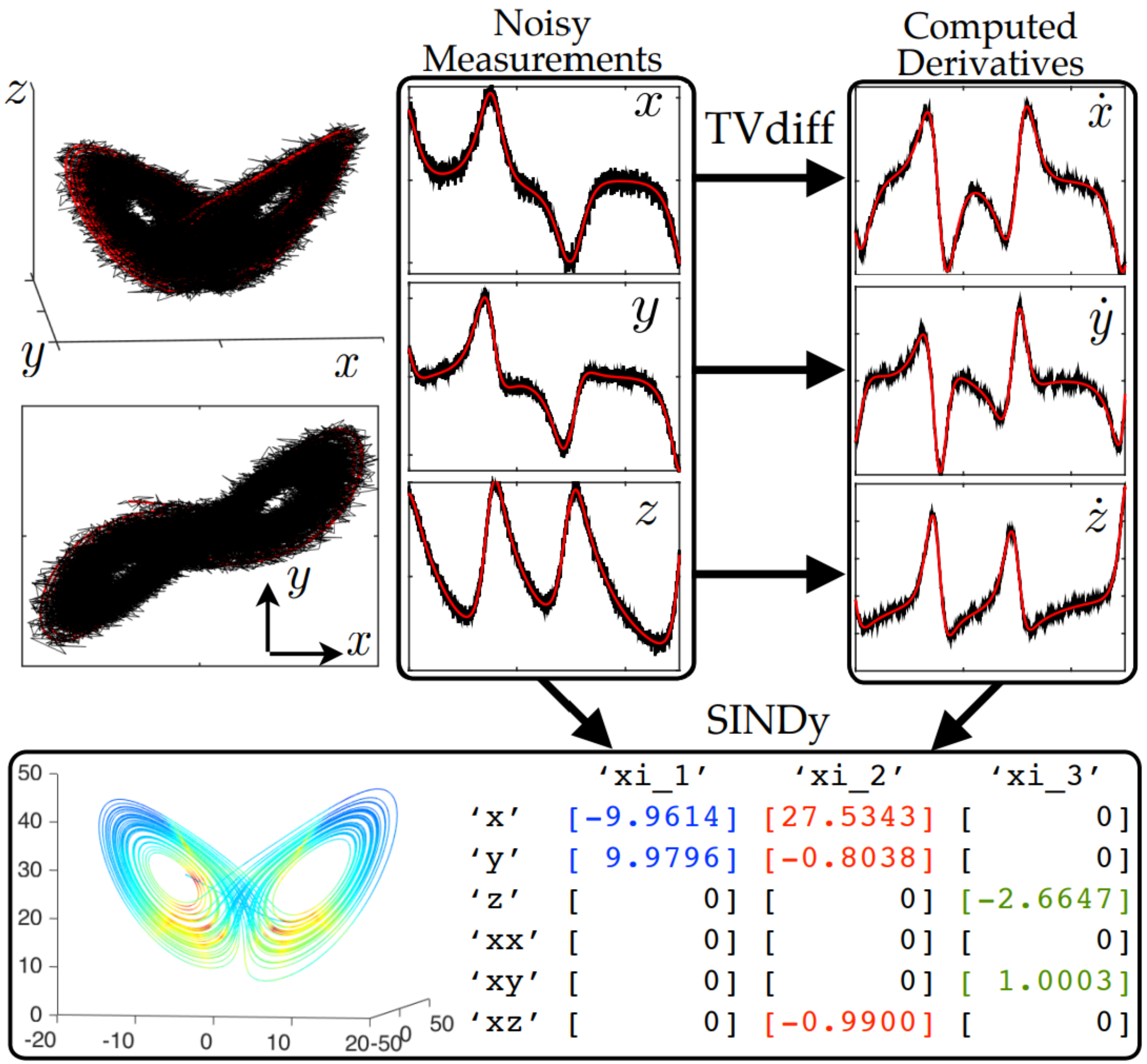
- No need to sample full attractor to identify correct model

Rate

- Higher rate improves identification
  - … or can reduces duration

However: clean data …

Data requirements – noise



Noisy measurements

- SINDy performance drops drastically
- Need denoising/filtering data
  - Low-pass filter
  - Total variation regularized derivative
    - Avoids noise amplification of finite-difference methods
  - PDE: polynomial interpolation

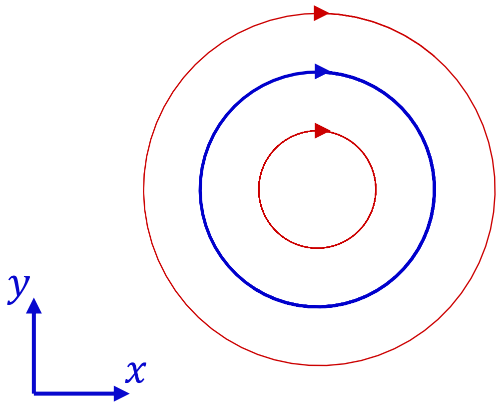
Noise robust SINDy

- Weak form & ensemble SINDy
  - *Discussed in next lecture*

SL Brunton, JL Proctor, JN Kutz (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. (supplementary material)

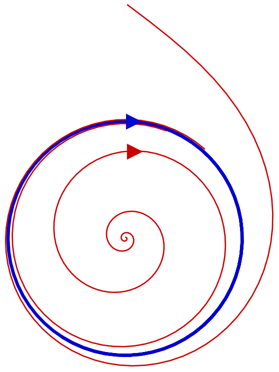
Data requirements – disambiguating models

Linear oscillator



$$\dot{x} = \omega y$$
$$\dot{y} = -\omega x$$

Cubic Hopf normal form



$$\dot{x} = \mu x + \omega y - Ax\left(x^2 + y^2\right)$$
$$\dot{y} = -\omega x + \mu y - Ay\left(x^2 + y^2\right)$$

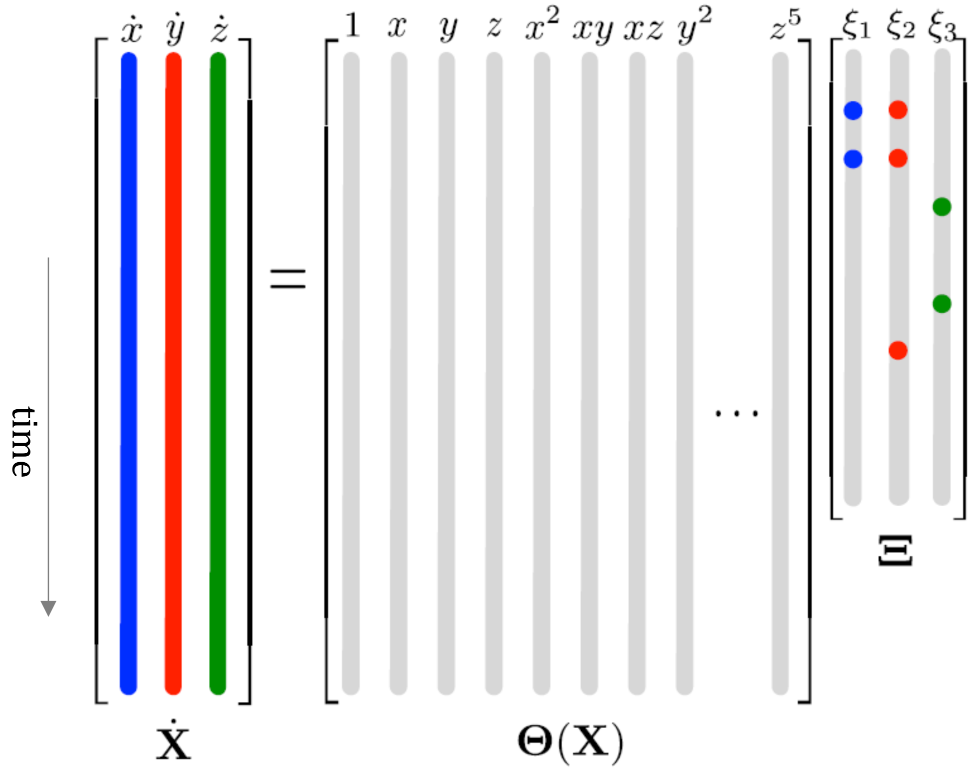
Multiple consistent models

- Collecting dynamical system data on attractor
  - Simplest model: linear oscillator
  - Other valid model: cubic Hopf normal form

    → Both models describe limit cycle behavior
  - **SINDy**: cubic and linear terms are parallel in $\boldsymbol{\Theta}(\mathbf{X})$ if we only sample data on circle
- Collect data from different experiments
  - Different initial conditions exciting different transient
  - Improves conditioning of library matrix $\boldsymbol{\Theta}(\mathbf{X})$

Library

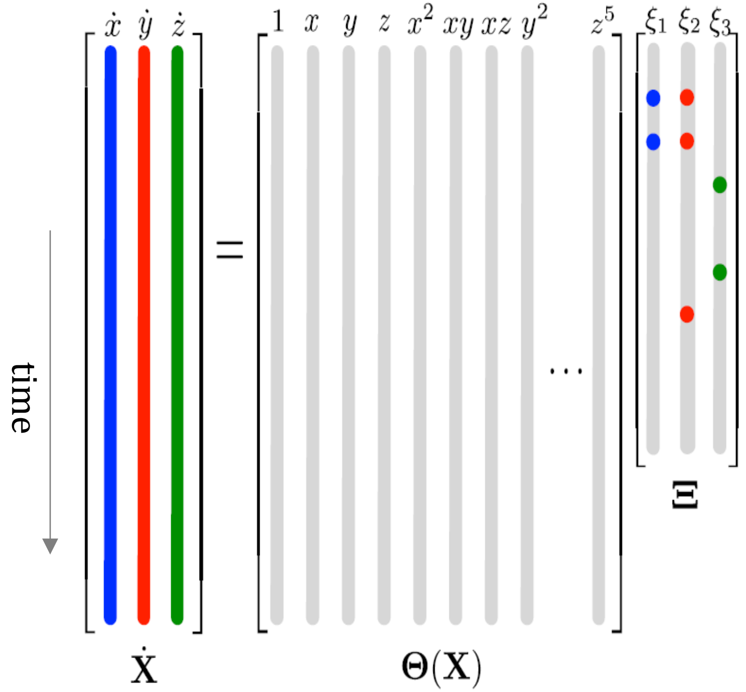
- **General approach**
- Rational functions
- Curse of dimensionality



Library – general approach

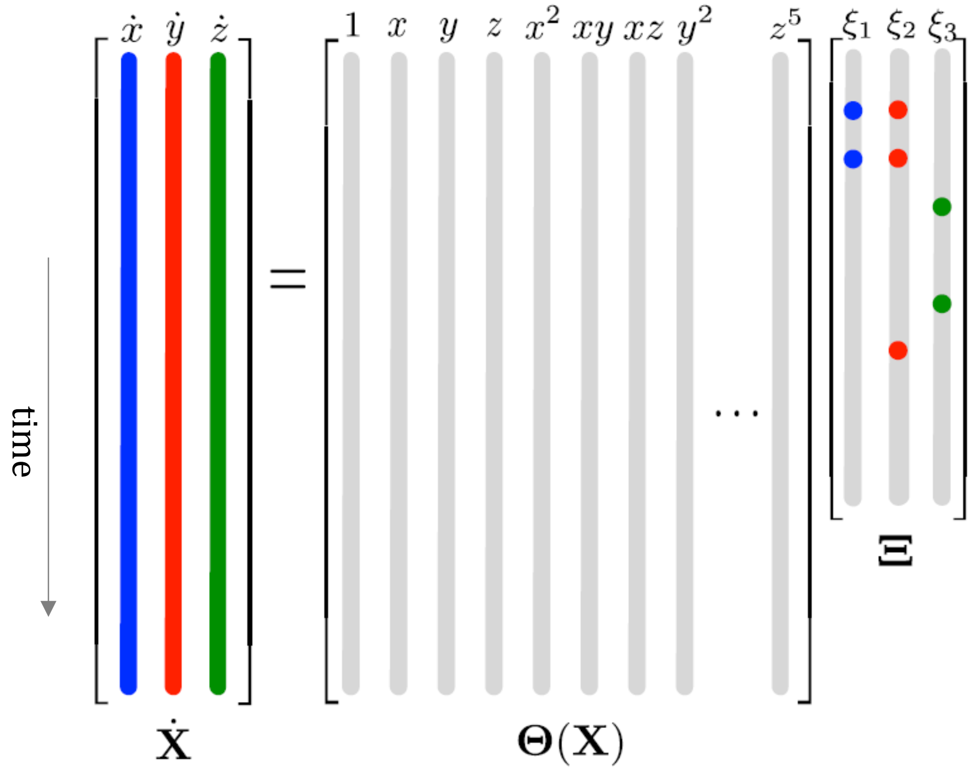
General approach to selecting the library

- Start with linear terms → DMD model
  - Check accuracy
    - error reconstructing $\dot{\mathbf{X}}$
    - model prediction error
- Increase order
  - Add quadratic, then higher order polynomials
- Trigonometric functions
- <u>Generally</u>: try small, isolated libraries first



Library

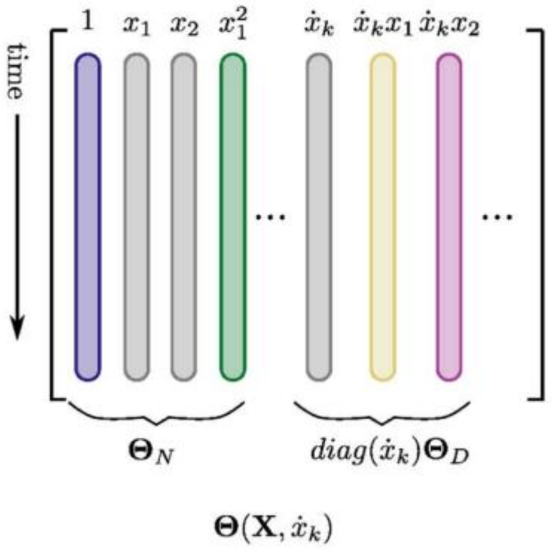
- General approach
- **Rational functions**
- Curse of dimensionality

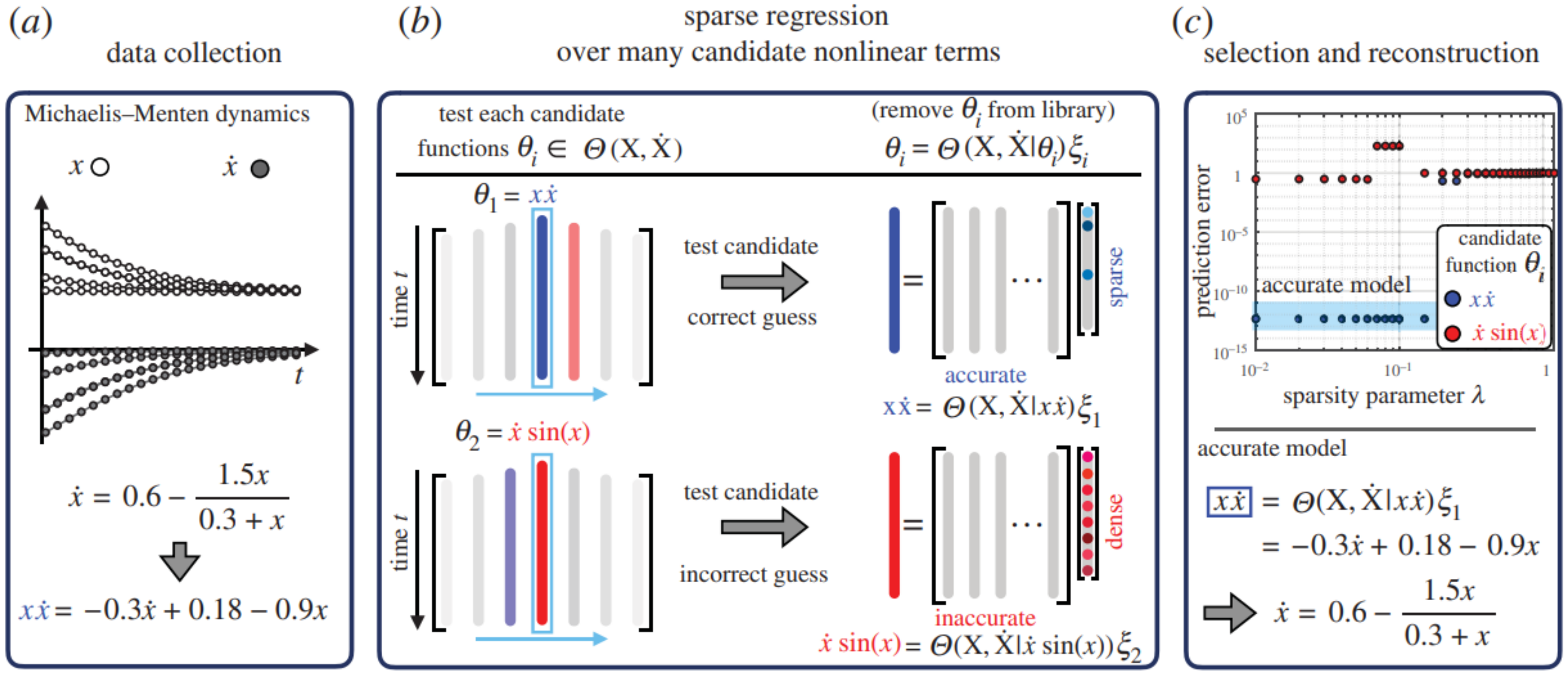


Library – rational functions

Extending SINDy: handle larger classes of dynamical systems

- Rational function: $\dot{x}_k = \dfrac{f_N(\mathbf{x})}{f_D(\mathbf{x})} \quad \rightarrow \quad f_N(\mathbf{x}) - f_D(\mathbf{x})\dot{x}_k = 0 \quad \rightarrow \quad \boldsymbol{\Theta}(\mathbf{X}, \dot{x}_k)\boldsymbol{\xi}_k = \mathbf{0}$
  - difficult to describe as a linear combination of library features
  - e.g. biological systems: Michaelis-Menten dynamics: $\dot{x} = 0.6 - \dfrac{1.5x}{0.3+x}$



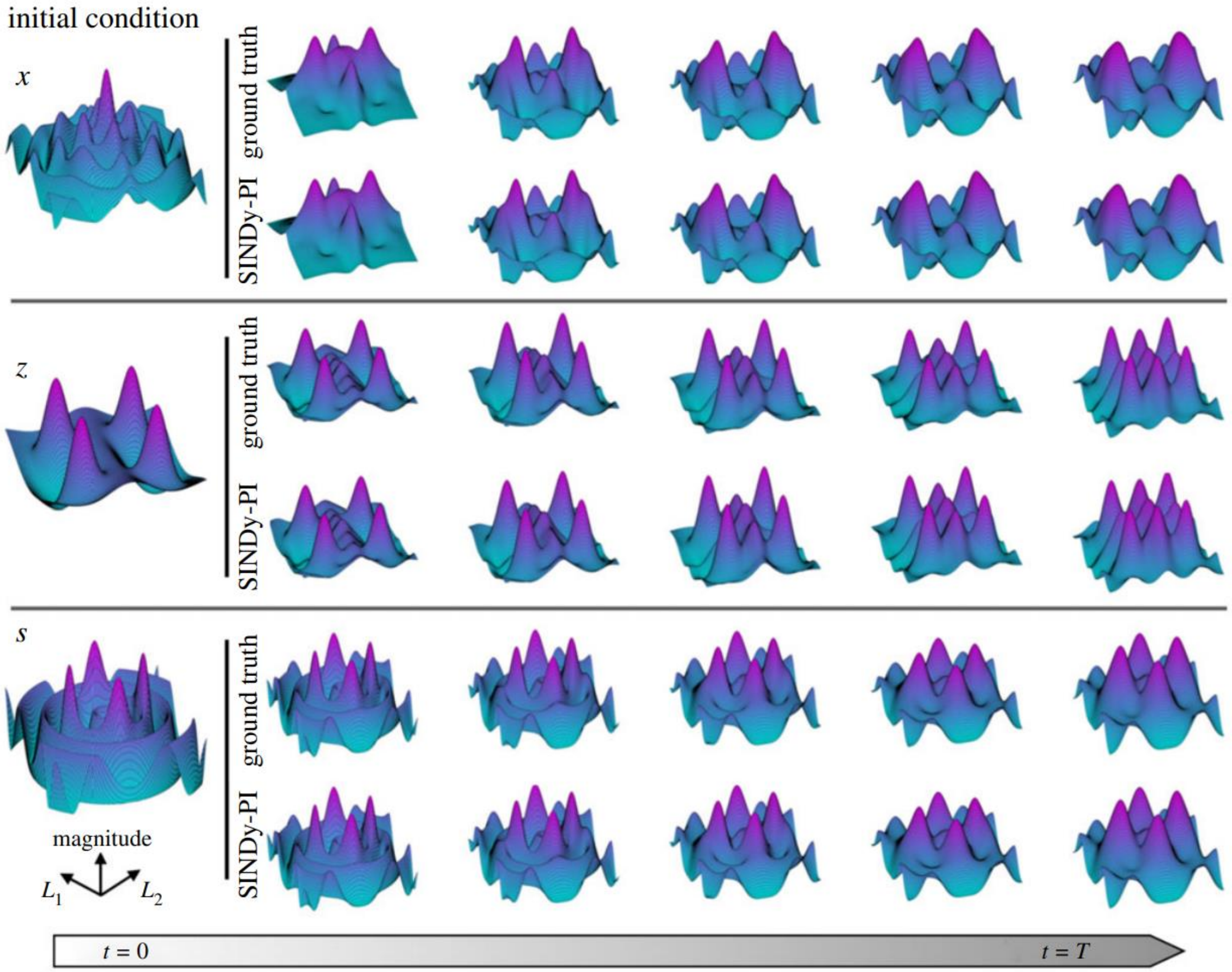


SINDy-PI algorithm

1. Test multiple possible left-hand sides (in parallel):
   - Move candidate terms to LHS
2. Calc model prediction error
3. Select best model:
   - Sparsity & accuracy

NM Mangan, SL Brunton, JL Proctor, JN Kutz (2016) Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics
K Kaheman, JN Kutz, SL Brunton (2020) SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics

Library – rational functions



Belousov–Zhabotinsky reaction

- 4 coupled PDE with rational nonlinearities

$$\frac{\partial x}{\partial \tau} = \frac{1}{\varepsilon}\left(\frac{fz(q-x)}{q+x} + x - x^2 - \beta x + s\right) + \frac{D_x}{D_u}\Delta x,$$

$$\frac{\partial z}{\partial \tau} = x - z - \alpha z + \gamma u + \frac{D_z}{D_u}\Delta z,$$

$$\frac{\partial s}{\partial \tau} = \frac{1}{\varepsilon_2}(\beta x - s + \chi u) + \frac{D_s}{D_u}\Delta s,$$

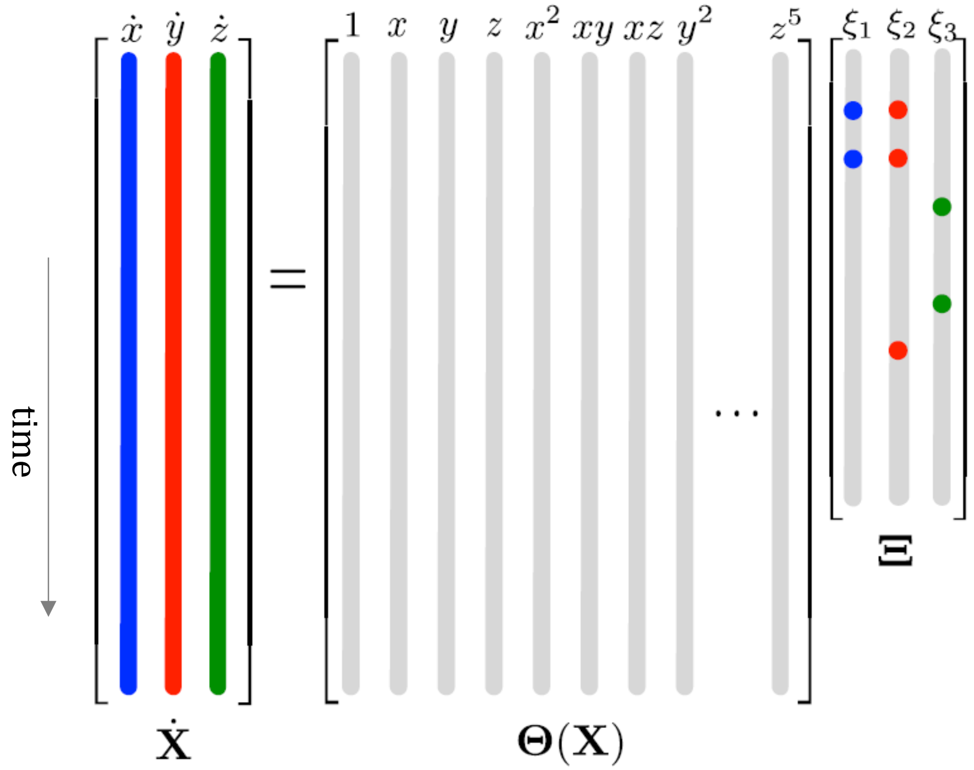
$$\frac{\partial u}{\partial \tau} = \frac{1}{\varepsilon_3}\left[\alpha z - \left(\gamma + \frac{\chi}{2}\right)u\right] + \frac{D_u}{D_u}\Delta u,$$

- SINDy-PI accurately identifies correct PDE
  - Not possible with standard SINDy
  - **MATLAB**: https://github.com/dynamicslab/SINDy-PI
  - **PySINDy**: interactive notebook

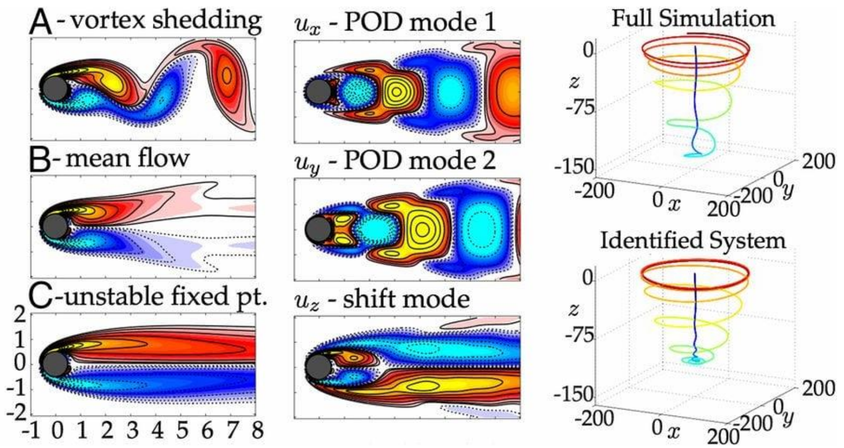
K Kaheman, JN Kutz, SL Brunton (2020) SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics

Library

- General approach
- Rational functions
- **Curse of dimensionality**



Library – curse of dimensionality

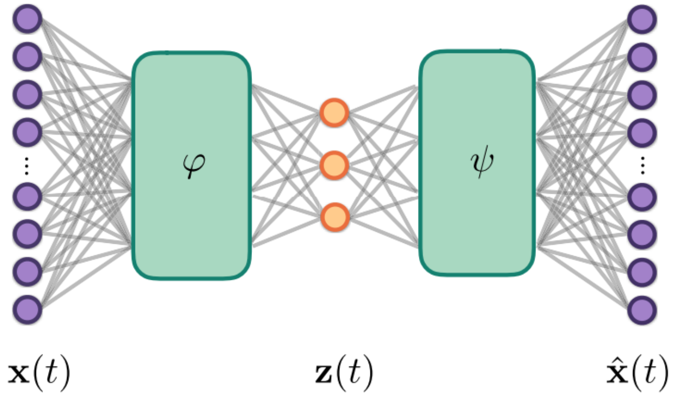


Challenge: Sparse regression quickly becomes intractable

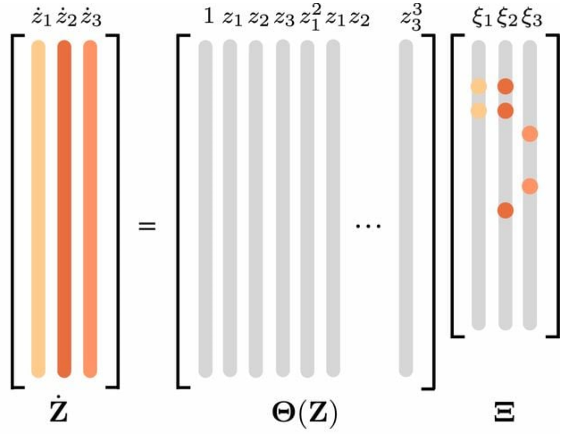
- e.g. large state dimension or large order of polynomials

→ **Reduce dimension of the state $\mathbf{x}$**

1. Dimensionality reduction:



- **SVD/PCA/POD**
  - Original SINDy paper 2016: identify dynamics of POD mode amplitudes
- **Autoencoders**: simultaneously discover coordinates and sparse models
  - Nonlinear generalization of PCA → can possibly further reduce state dimension
  - Optimizer jointly minimizes autoencoder reconstruction loss and SINDy loss
    - Autoencoder loss: $\|\mathbf{x} - \psi(\phi(\mathbf{x}))\|_2^2$
    - SINDy loss: $\|\dot{\mathbf{z}} - \mathbf{\Theta}(\mathbf{z})\mathbf{\Xi}\|_2^2 + \lambda\|\mathbf{\Xi}\|_0$



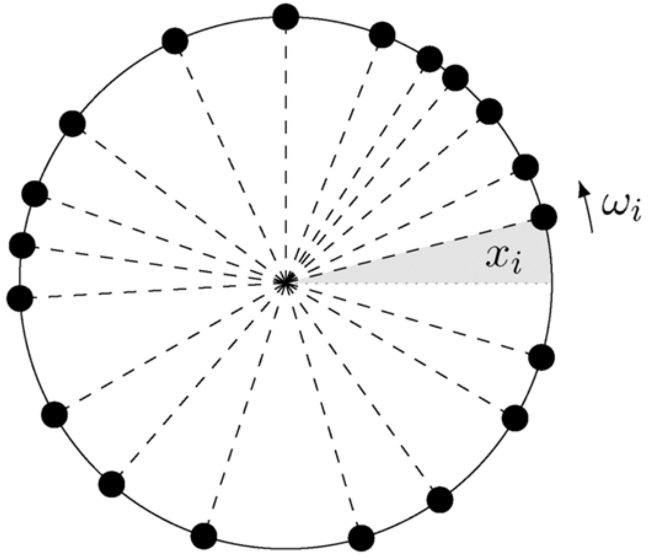
- **Paper**: K Champion, B Lusch, JN Kutz, SL Brunton (2019) Data-driven discovery of coordinates and governing equations.

Library – curse of dimensionality

Challenge: Sparse regression quickly becomes intractable

- e.g. large state dimension or large order of polynomials

→ **Different method to compute least squares (pseudo inverse)**



2. Tensor SINDy

- Generalize SINDy library approach to include tensor train formulations
  - Using low-rank tensor decomposition to learn high dimensional dynamics
- **Example**: Kuramoto model → 100 coupled oscillators on a ring
  - Significantly reduces memory consumption and computational cost
- **Paper**: P Gelß et al (2019) Multidimensional Approximation of Nonlinear Dynamical Systems.

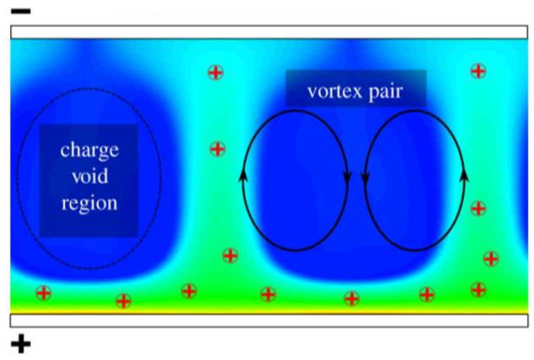
Library – curse of dimensionality

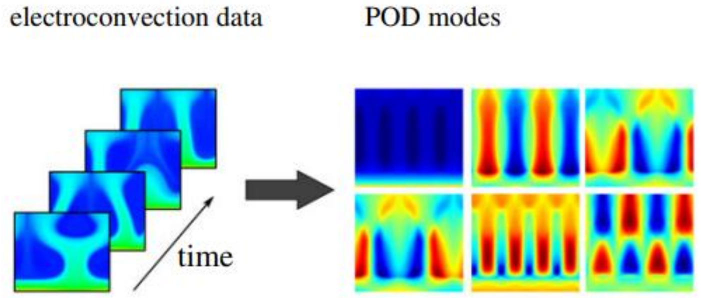
3. Include prior knowledge to constrain the library

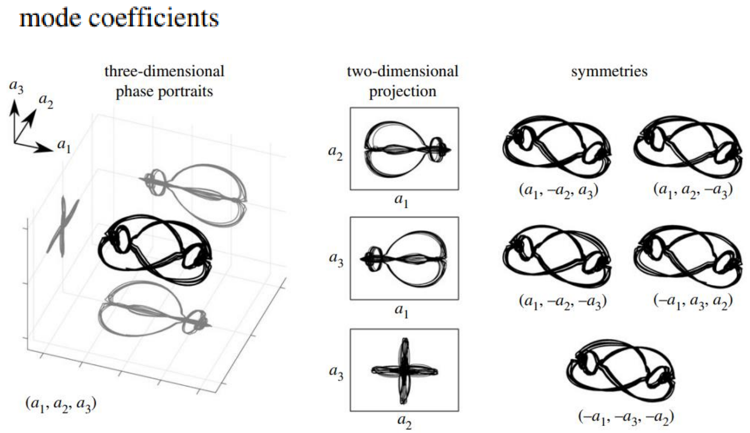
- Physical symmetries, energy-preserving quadratic nonlinearities, …
- **Constrained SINDy:** $\mathrm{argmin}_{\Xi} \left\|\dot{\mathbf{X}} - \boldsymbol{\Theta}(\mathbf{X})\Xi\right\|_2^2 + \lambda\|\Xi\|_0$ subject to $\mathbf{C}\boldsymbol{\xi} = \boldsymbol{d}$
  - $\xi = \Xi(:)$**,** vectorized form of coefficient matrix
  - With STLS: constraints are imposed via Lagrange multipliers → penalized least squares
- **Electroconvection:** Dielectric fluid between parallel electrodes under strong unipolar injection
  - Unsteady ionic convection leads to electric field variation and consequently the unsteady flow patterns
- **Trajectories exhibit symmetries** → system invariant with respect to some transformations:
  - $\boxed{[a_1, a_2, a_3] \leftrightarrow [a_1, -a_2, a_3] \leftrightarrow [a_1, a_2, -a_3] \leftrightarrow [a_1, -a_2, -a_3]} \leftrightarrow [-a_1, a_2, a_3] \leftrightarrow [-a_1, -a_2, -a_3]$
  - Symmetries constrain library: e.g. $\dot{a}_1$ dynamics invariant to switching sign of a_2 and/or a_3.
    - → $\dot{a}_1$ library terms with odd powers of either a_2 or a_3 must vanish
    - → these constraints reduce library size and “simplify” problem:

	a_1	a_2	a_3	a_1^2	a_1a_2	a_1a_3	a_2^2	a_2a_3	a_3^2	a_1^3	$a_1^2a_2$	$a_1^2a_3$	$a_1a_2^2$	$a_1a_2a_3$	$a_1a_3^2$	a_2^3	$a_2^2a_3$	$a_2a_3^2$	a_3^3
$\dot{a}_1$	ξ_1						ξ_2^*		$-\xi_2$	ξ_3			ξ_4		ξ_4				
$\dot{a}_2$		ξ_5			$-\xi_2^*$						ξ_6					ξ_7		ξ_8	
$\dot{a}_3$			ξ_5			ξ_2						ξ_6					ξ_8		ξ_7

Charge density from electroconvection simulation



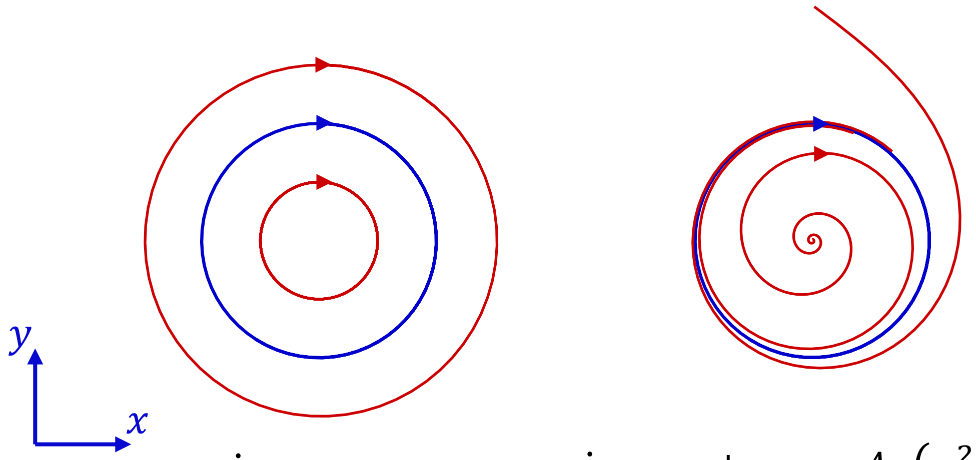




- **1st paper on constrained SINDy**: JC Loiseau, SL Brunton (2018) Constrained sparse Galerkin regression.
- **Electroconvection**: Y Guan et al(2021) Sparse nonlinear models of chaotic electroconvection.
- **SINDy models enforcing stable dynamics**: A Kaptanoglu et al (2021) Promoting global stability in data-driven models of quadratic nonlinear dynamics. → GitHub

Tutorial summary

- **Data requirements**
  - Sampling duration & rate
  - Noise
  - Disambiguating multiple consistent models



$$\dot{x} = \omega y$$
$$\dot{y} = -\omega x$$

$$\dot{x} = \mu x + \omega y - Ax(x^2 + y^2)$$
$$\dot{y} = -\omega x + \mu y - Ay(x^2 + y^2)$$

- **Library**
  - Rational functions
  - Curse of dimensionality