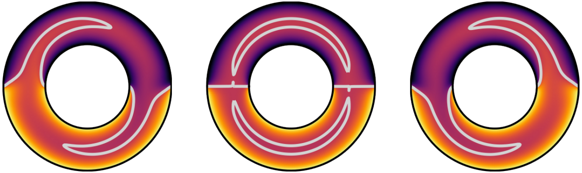


Dimensionality reduction and system identification in fluid dynamics

Jean-Christophe Loiseau

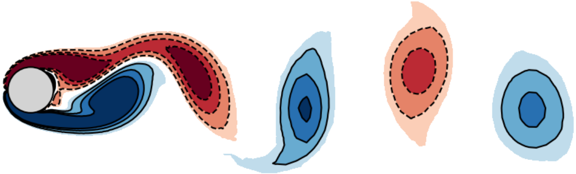
DynFluid, Arts & Métiers

Who am I?

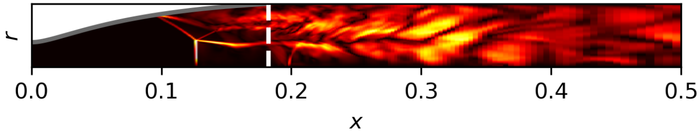
- Maître de Conférences in Fluid Dynamics (but mainly teaching Applied Mathematics).
- Machine-Learning enthusiast.
- Particular emphasis on the application of ML to engineering systems.
- Excited by data-efficient models with some kind of guarantees of optimality.

A gallery of fluid examples

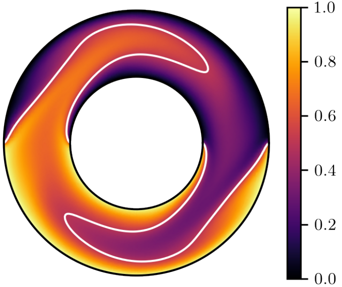
Illustrative examples



Aerodynamics



Rocket Science (litterraly!)

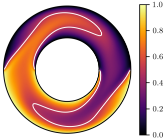


Heat exchange

A gallery of fluid examples

Machine learning for physical systems

High-dimensional system

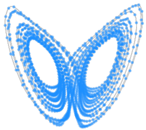


Forecasting,
Feedback control

Dimensionality
reduction

$$
\begin{aligned}
\dot{x} &= \alpha_1 x + \alpha_2 y \\
\dot{y} &= \beta_1 x + \beta_2 y + \beta_3 xz \\
\dot{z} &= \gamma_1 + \gamma_2 z + \gamma_3 xy
\end{aligned}
$$

Dynamical model



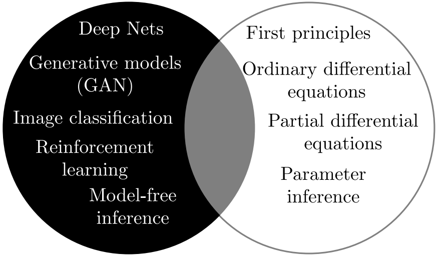
Latent space

System Identification

A gallery of fluid examples

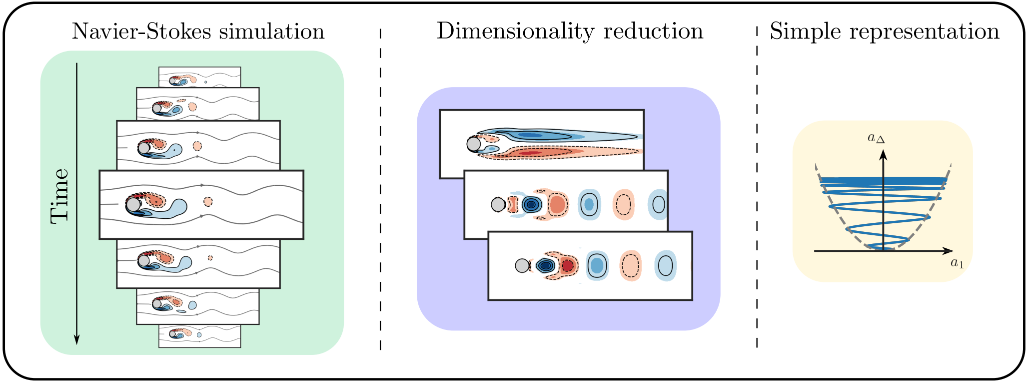
Low-order models

- Models can come in a lot of different flavors.
- Wiener proposed the following classification
  - ↪ **Black box:** input-output data.
  - ↪ **White box:** Input-output + known model.
  - ↪ **Gray box:** Input-output + partial knowledge of the model / inductive biases.
- In physics, we are often in the third class.
  - ↪ Exact form of the equations may be unknown.
  - ↪ Knowledge of symmetries or invariants.



A gallery of fluid examples

A prototypical example



A gallery of fluid examples

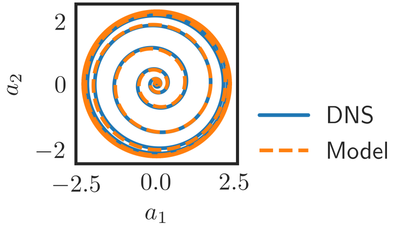
Low-order deterministic models

Aim : Identify an interpretable and physically-consistent low-order model.

- Using *SINDy*, the following model can be identified.

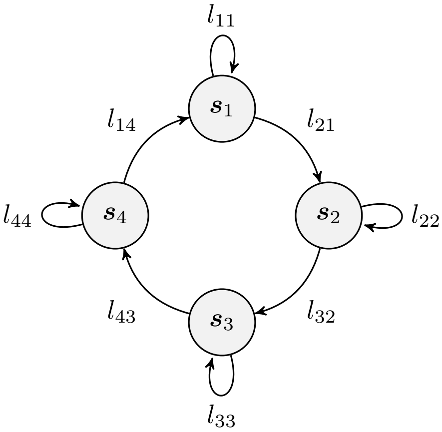
$$\dot{a}_1 = \sigma a_1 - \omega a_2 - \alpha(a_1^2 + a_2^2)a_1 - \beta(a_1^2 + a_2^2)a_2$$
$$\dot{a}_2 = \omega a_1 + \sigma a_2 + \beta(a_1^2 + a_2^2)a_1 - \alpha(a_1^2 + a_2^2)a_1.$$

- **Interpretability :** Normal form of a supercritical Andronov-Poincaré-Hopf bifurcation.
- **Physical-consistency :** By design.



A gallery of fluid examples

Low-order probabilistic models



Aim : Identify a coarse-grain probabilistic low-order model.

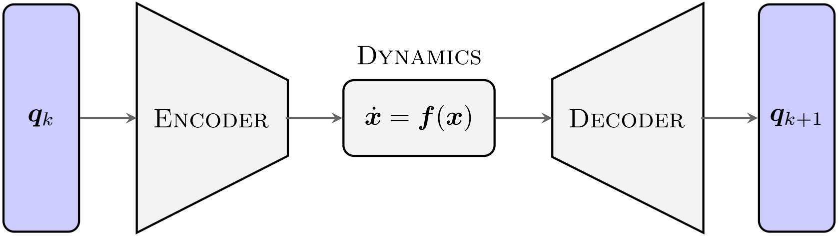
- Using clustering and symbolic dynamics, a discrete-time probabilistic model can be identified.

$$\boldsymbol{p}_{k+1} = \boldsymbol{L}\boldsymbol{p}_k.$$

- **Interpretability :** $\boldsymbol{L}$ describes the transition probability between discrete states $\boldsymbol{s}_i$.
- Dynamics are encoded as a probabilistic finite state machine.

Low-order modeling

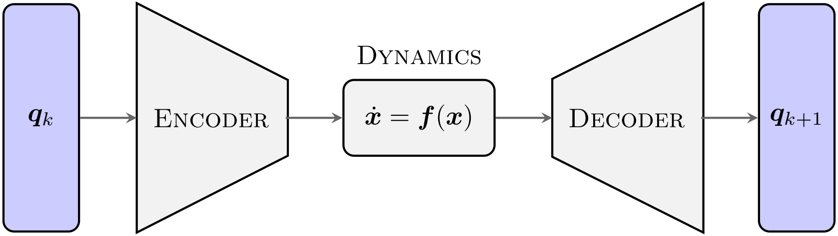
Ideal setup



Ideal setup : Learn jointly the encoding from the high-dimensional space to the low-dimensional one and the dynamical model within this subspace.

Low-order modeling

In practice



In practice : Learning jointly the encoder/decoder and the dynamical model is complicated. In practice, they are learned sequentially (i.e. first encoder/decoder and then the dynamical model).

Our setup for today

2D Thermosiphon

- Two-dimensional flow governed by the incompressible Navier-Stokes equations

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p + \mathrm{Pr}\nabla^2 \boldsymbol{u} + \mathrm{Ra}\ \mathrm{Pr}\ \boldsymbol{\theta} \boldsymbol{e}_y$$
$$\frac{\partial \boldsymbol{\theta}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{\theta} = \nabla^2 \boldsymbol{\theta}.$$

- Ra and Pr are parameters defining our problem.
  - ↪ Ra is set to $Ra = 17\ 000$.
  - ↪ Pr is set to $Pr = 5$.

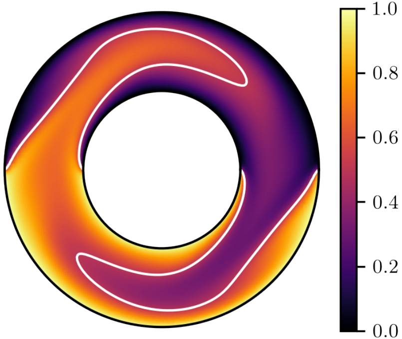
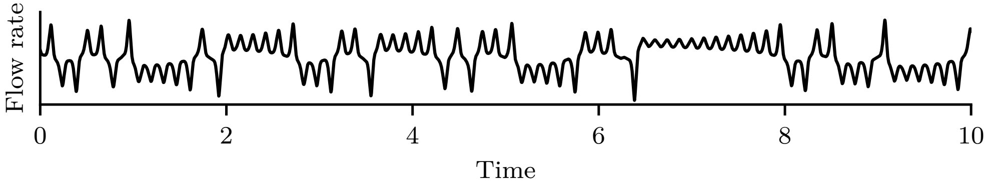


Fig: Chaotic thermosyphon temp. field.

Our setup for today

Chaotic dynamics



- Time-evolution of the cross-sectional flow rate is indicative of Lorenz-like chaotic dynamics.
  ↪ "Random" switching between clockwise and anti-clockwise rotation.

- **Hypothesis**: These dynamics can be captured by a low-order model.

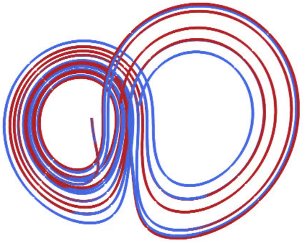
Our setup for today

A quick detour

- Most notorious chaotic dynamical system proposed by Edward Lorenz in 1963. It reads

$$\begin{aligned}\dot{x} &= \sigma \left(y - x \right) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= xy - \beta z.\end{aligned}$$

- It can be derived (under certain conditions) directly from the Navier-Stokes equations.

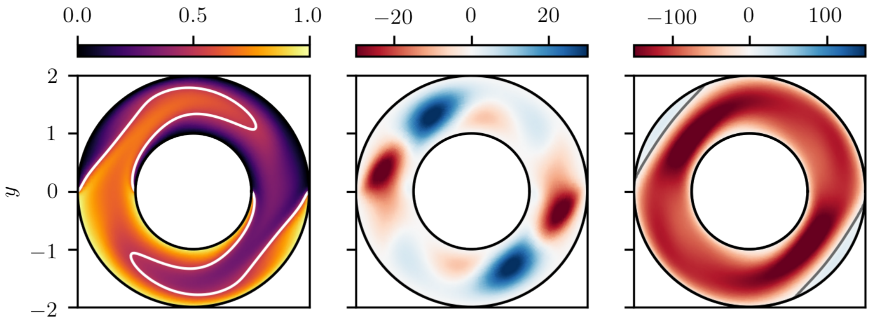


Objectives

- What physical properties should our reduced-order model have?
  - ↪ Analyze the physics prior to modeling.
- How to obtain a good low-dimensional embedding?
  - ↪ Dimensionality reduction.
- Can we identify the equations governing the dynamics in the embedded space?
  - ↪ System identification.
  - ↪ How to enforce the physical constraints?

Chaotic thermosyphon

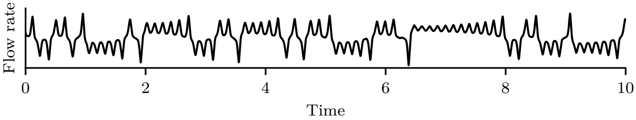
Characterizing its dynamics



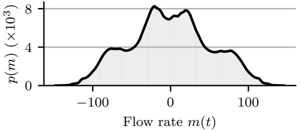
(a) Temperature (b) Radial velocity (c) Azimuthal velocity

Chaotic thermosyphon

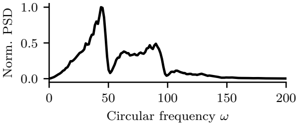
Characterizing its dynamics



(a) Flow rate $m(t)$



(b) Empirical PDF

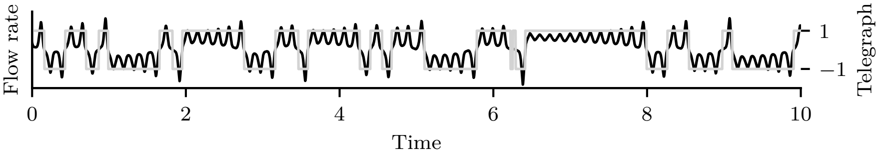


(c) PSD of $\dot{m}(t)$

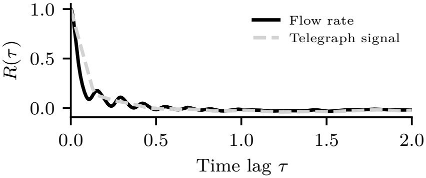
- Clockwise and counter-clockwise rotations are equally likely.
- Dominant time-scale associated w/ the oscillation in one direction.
- Continuous spectrum due to the chaotic nature of the system.

Chaotic thermosyphon

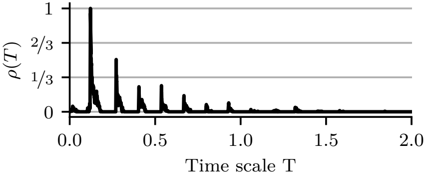
Characterizing its dynamics



(d) Flow rate $m(t)$



(e) Autocorr. function

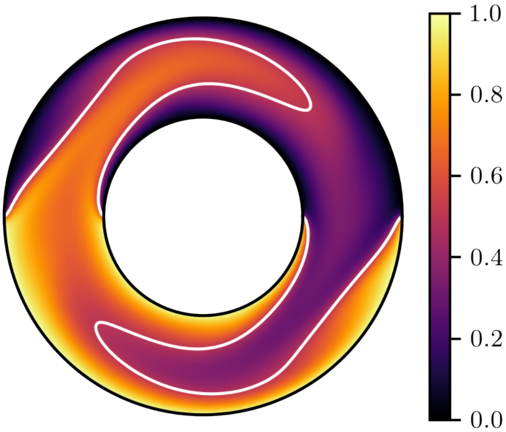


(f) Time-scale dist.

Part I

Dimensionality reduction

- Problem formulation:
  - ↪ DMD as a Reduced Rank Regression
  - ↪ Linear autoencoder + linear dynamics
- Practical usage:
  - ↪ Extracting patterns from our flow
  - ↪ Low-dimensional embedding



Dynamic Mode Decomposition

Problem formulation

- Given a discrete-time system $\boldsymbol{x}_{k+1} = \boldsymbol{f}(\boldsymbol{x}_k)$, DMD aims to find a (low-rank) linear operator $\boldsymbol{A}$ such that

$$\underset{\boldsymbol{A}}{\text{minimize}} \ \sum_{k=1}^{n} \|\boldsymbol{f}(\boldsymbol{x}_k) - \boldsymbol{A}\boldsymbol{x}_k\|_2^2$$
$$\text{subject to rank } \boldsymbol{A} = r.$$

- This problem belongs to the class of *Reduced-Rank Regression* problems.

Reduced-Rank Regression for the Multivariate Linear Model

ALAN JULIAN IZENMAN

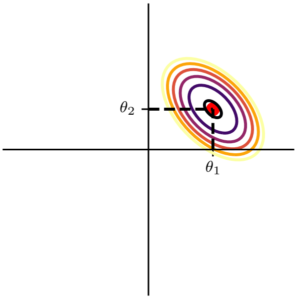
Department of Statistics, Tel Aviv University, Israel

Communicated by P. R. Krishnaiah

The problem of estimating the regression coefficient matrix having known (reduced) rank for the multivariate linear model when both sets of variates are jointly stochastic is discussed. We show that this problem is related to the problem of deciding how many principal components or pairs of canonical variates to use in any practical situation. Under the assumption of joint normality of the two sets of variates, we give the asymptotic (large-sample) distributions of the various estimated reduced-rank regression coefficient matrices that are of interest. Approximate confidence bounds on the elements of these matrices are then suggested using either the appropriate asymptotic expressions or the jackknife technique.

Dynamic Mode Decomposition

Matrix formulation and unconstrained solution



- Problem can be recast in matrix form as

$$\underset{\boldsymbol{A}}{\text{minimize}} \ \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}\|_F^2$$
$$\text{subject to rank } \boldsymbol{A} = r,$$

where $\boldsymbol{Y} = \boldsymbol{X}_{k+1}$ and $\boldsymbol{X} = \boldsymbol{X}_k$.

- Unconstrained solution is simply the least-squares solution

$$\boldsymbol{A} = \boldsymbol{C}_{\boldsymbol{yx}}\boldsymbol{C}_{\boldsymbol{xx}}^{-1},$$

with $\boldsymbol{C}_{\boldsymbol{yx}} = \boldsymbol{Y}\boldsymbol{X}^H$ and $\boldsymbol{C}_{\boldsymbol{xx}} = \boldsymbol{X}\boldsymbol{X}^H$.

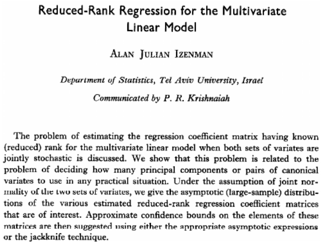
Dynamic Mode Decomposition

Rank-constrained solution

- Introducing the low-rank factorization $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{Q}^H$, the minimization problem reads

$$
\begin{aligned}
\underset{\boldsymbol{P},\boldsymbol{Q}}{\text{minimize}} \ & \|\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^H\boldsymbol{X}\|_F^2 \\
\text{subject to } & \operatorname{rank} \boldsymbol{P} = \operatorname{rank} \boldsymbol{Q} = r \\
& \boldsymbol{P}^H\boldsymbol{P} = \boldsymbol{I}.
\end{aligned}
$$

- $\boldsymbol{P}$ is a basis for the column-span of $\boldsymbol{A}$ while $\boldsymbol{Q}$ is a basis for its row-span. These are in general different.



Reduced-Rank Regression for the Multivariate Linear Model

Alan Julian Izenman

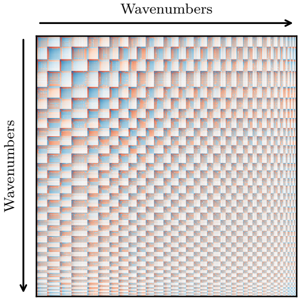
Department of Statistics, Tel Aviv University, Israel

Communicated by P. R. Krishnaiah

The problem of estimating the regression coefficient matrix having known (reduced) rank for the multivariate linear model when both sets of variates are jointly stochastic is discussed. We show that this problem is related to the problem of deciding how many principal components or pairs of canonical variates to use in any practical situation. Under the assumption of joint normality of the two sets of variates, we give the asymptotic (large-sample) distributions of the various estimated reduced-rank regression coefficient matrices that are of interest. Approximate confidence bounds on the elements of these matrices are then suggested using either the appropriate asymptotic expressions or the jackknife technique.

Dynamic Mode Decomposition

Rank-constrained solution



- Our minimization problem is equivalent to

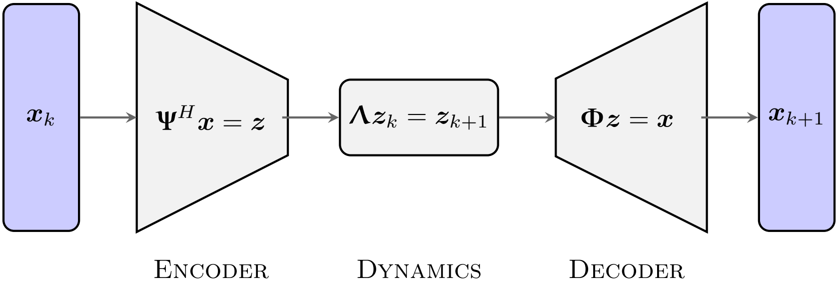
$$\begin{aligned} \underset{\boldsymbol{P}}{\text{maximize}} \; & \operatorname{Tr}\left(\boldsymbol{P}^H \boldsymbol{C}_{\boldsymbol{yx}} \boldsymbol{C}_{\boldsymbol{xx}}^{-1} \boldsymbol{C}_{\boldsymbol{xy}} \boldsymbol{P}\right) \\ \text{subject to} \; & \operatorname{rank} \boldsymbol{P} = r \\ & \boldsymbol{P}^H \boldsymbol{P} = \boldsymbol{I}. \end{aligned}$$

- Optimal solution is given by the first r eigenvectors of the symmetric positive-definite matrix $\boldsymbol{C}_{\boldsymbol{yx}} \boldsymbol{C}_{\boldsymbol{xx}}^{-1} \boldsymbol{C}_{\boldsymbol{xy}}$.
- $\boldsymbol{Q}$ is solution to $\boldsymbol{C}_{\boldsymbol{xx}} \boldsymbol{Q} = \boldsymbol{C}_{\boldsymbol{xy}} \boldsymbol{P}$.

Dynamic Mode Decomposition

Dimensionality reduction and linear model

- Once $\boldsymbol{P}$ and $\boldsymbol{Q}$ are found, $\boldsymbol{A}$ may be factorized as $\boldsymbol{A} = \boldsymbol{\Phi\Lambda\Psi}^H$.

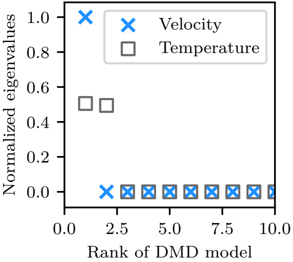


Interpretation : ℓ_2-optimal combination of linear autoencoder for dimensionality reduction and linear time-invariant model for the latent space (one-step ahead) dynamics.

Dynamic Mode Decomposition

In practice

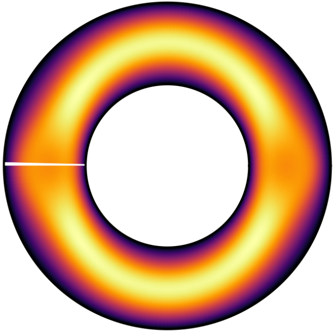
- Sampling period is heuristically chosen to be $\tau_0/20$.
- DMD models are fitted separately for the velocity and temperature fields based on 20 000 snapshots.



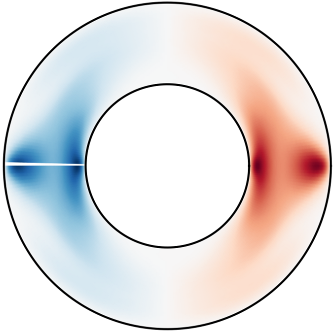
Eigenspectrum of $\boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy}$.

Dynamic Mode Decomposition

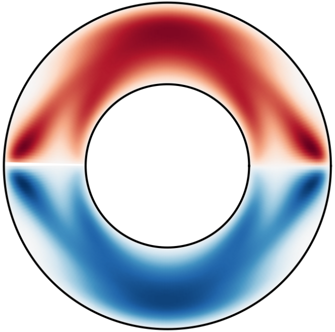
In practice



Leading DMD mode for the velocity.



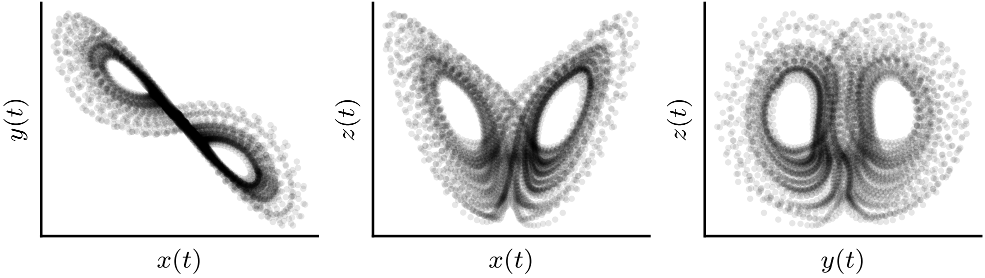
First DMD mode for the temperature.

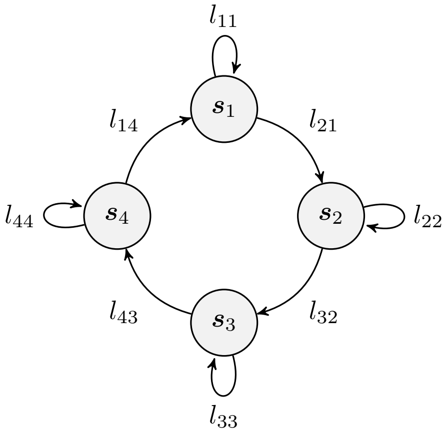


Second DMD mode for the temperature.

Dynamic Mode Decomposition

In practice





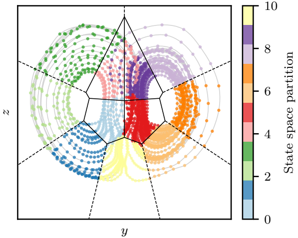
Part II

Statistics of chaotic

- Problem formulation:
  - ↪ State space partition
  - ↪ Cluster-based Reduced Order Modeling
- Practical usage:
  - ↪ Probabilistic model
  - ↪ Statistics of chaos

Statistics of chaotic

Overview



Cluster-based Reduced-Order Modeling

- **Aim:** Coarse-gain probabilistic model of the dynamics
  - ↪ Linear model for the probability density function.
  - ↪ Statistical properties of the system.
- **How:** From continuous to symbolic dynamics.
  - ↪ State-space partitionning using k-means.
  - ↪ Maximum likelihood estimation of the transition probability from cluster c_i to cluster c_j.

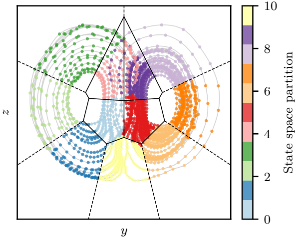
Statistics of chaotic

State-space partitionning

- Given a set of n observations $(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$, partition the data into k sets $\boldsymbol{C} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \cdots, \boldsymbol{C}_k\}$.
- To do so, k-means aims to minimize the *within-cluster sum of squares*

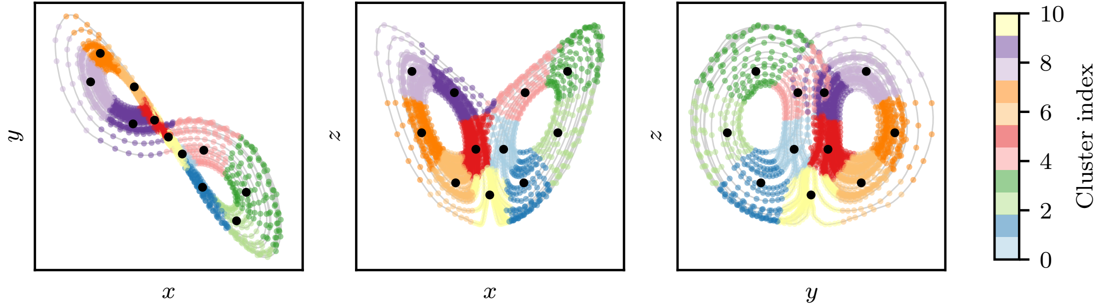
$$\underset{\boldsymbol{C}}{\text{minimize}} \sum_{i=1}^{k} \sum_{\boldsymbol{x} \in \boldsymbol{C}_i} \|\boldsymbol{x} - \boldsymbol{c}_i\|^2$$

where $\boldsymbol{c}_i$ is the centroid of the i[th] cluster $\boldsymbol{C}_i$.



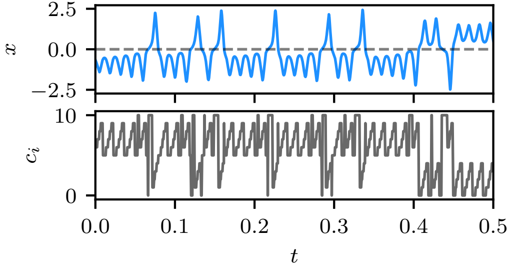
Statistics of chaotic

State-space partitionning



Statistics of chaotic

Symbolic dynamics



- Symbolic dynamics are obtained by looking at the evolution of the cluster index $c(t)$.
- These dynamics can easily be modeled using a Markov chain.

Statistics of chaotic

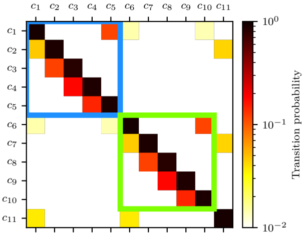
Markov chain model

- Using MLE, a Markov chain model

$$\boldsymbol{p}_{k+1} = \boldsymbol{L}\boldsymbol{p}_k$$

  is identified where the i$^{\text{th}}$ entry of $\boldsymbol{p}_k$ describes the probability of being in cluster $\boldsymbol{C}_i$ at time $t = k\Delta T$.

- l_{ij} encodes the transition probability from cluster $\boldsymbol{C}_j$ to cluster $\boldsymbol{C}_i$ after one step.

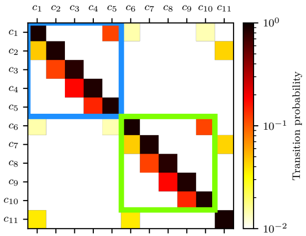


Transition matrix $\boldsymbol{L}$.

Statistics of chaotic

Markov chain model

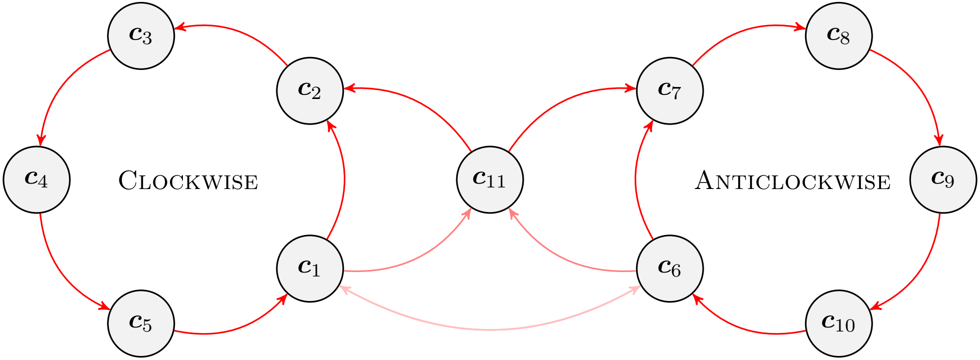
- Existence of two "mega-clusters" corresponding to the clockwise and counter-clockwise rotation.
- When reaching the last cluster, the system to randomly switch from one side of the attractor to the other.



Transition matrix $\boldsymbol{L}$.

Statistics of chaotic

Markov chain model

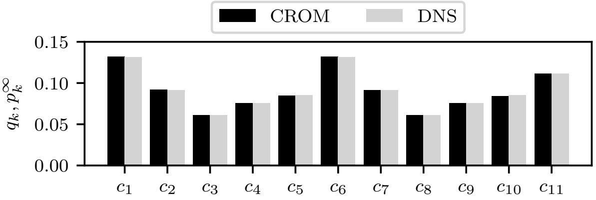


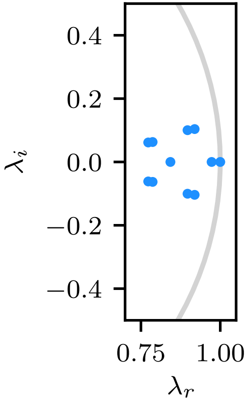
Statistics of chaotic

Statistical analysis

- Eigenspectrum of $\boldsymbol{L}$ encodes the diffusion of the probability density function.

- Eigenvector $\boldsymbol{p}^\infty$ associated to $\lambda_1 = 1$ characterizes the invariant measure (i.e. the long-time/ensemble average distribution of $\boldsymbol{p}$).

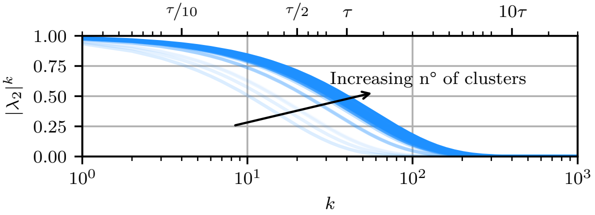


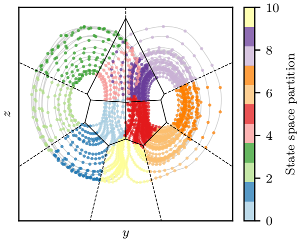


Statistics of chaotic

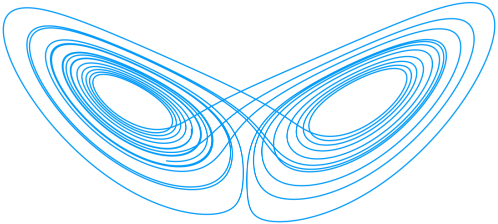
Statistical analysis

- How fast the p.d.f. diffuses is related to the ratio $|\lambda_2|/|\lambda_1|$.
  - ↪ The larger it is, the faster the p.d.f. converges to its invariant measure.
- $|\lambda_2|^k$ provides an estimate of how far in time we can predict.
  - ↪ Our prediction horizon is only one characteristic time-scale.





Part III



Sparse Identification of Nonlinear Dynamics

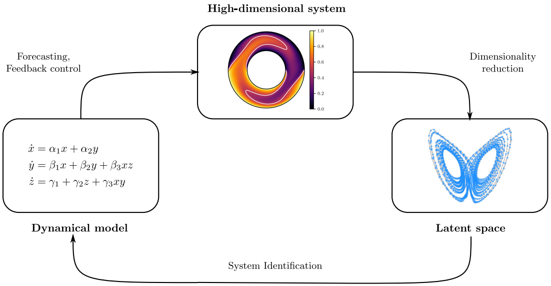
- Problem formulation:
  - ↪ Physical constraints
  - ↪ Optimization problem
  - ↪ Greedy algorithms and convex relaxations
- Application to the chaotic thermosyphon:
  - ↪ A Lorenz-like system
  - ↪ Cross-validations

Identifying a dynamical system

Overview

Sparse Identification of Nonlinear Dynamics

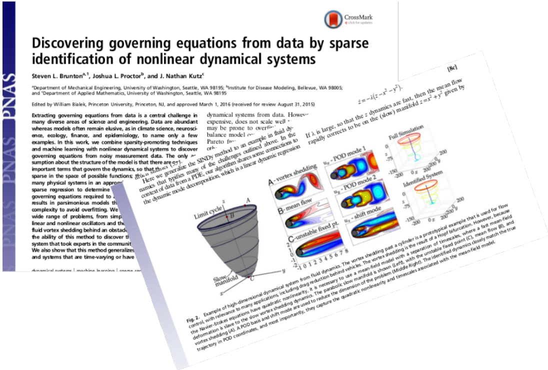
- **Aim:** Identify a low-order nonlinear dynamical system
  - ↪ Inherently interpretable set of ODE.
  - ↪ Fast evaluation for forecasting.
- **How:** Sparse least-squares regression
  - ↪ Constraints for physical consistency.
  - ↪ Greedy algorithms/convex relaxations.



Sparse Identification of Nonlinear Dynamics

Overview

- Brunton *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 2015.
- Relies on a dictionnary of pre-defined functions and sparsity-promoting regression.
- Quite promissing and versatile framework.

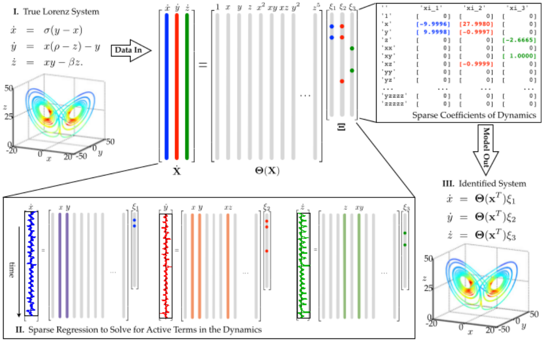


Discovering governing equations from data by sparse identification of nonlinear dynamical systems

Steven L. Brunton[a,1], Joshua L. Proctor[b], and J. Nathan Kutz[c]

Sparse Identification of Nonlinear Dynamics

A combinatorial problem



- Given $\boldsymbol{\Theta}(\boldsymbol{x})$, one aims to solve

$$\underset{\boldsymbol{\xi}}{\text{minimize}} \ \text{card}\ (\boldsymbol{\xi})$$
$$\text{subject to} \ \|\boldsymbol{\Theta}(\boldsymbol{x})\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2 \leq \sigma.$$

- Rapidly intractable combinatorial problem.
- Convex relaxation and/or greedy algorithms needed.

Sparse Identification of Nonlinear Dynamics

Problem formulation: general form

- Navier-Stokes are nonlinear PDEs with quadratic nonlinearities so the dictionnary is chosen as

$$\boldsymbol{\Theta}(\boldsymbol{x}) = \begin{bmatrix} 1 & x & y & z & x^2 & xy & xz & y^2 & yz & z^2 \end{bmatrix}.$$

- The yet-unknown model takes the following general form

$$\begin{aligned}
\dot{x} &= a_0 + a_1 x + a_2 y + a_3 z + a_4 x^2 + a_5 xy + a_6 xz + a_7 y^2 + a_8 yz + a_9 z^2 \\
\dot{y} &= b_0 + b_1 x + b_2 y + b_3 z + b_4 x^2 + b_5 xy + b_6 xz + b_7 y^2 + b_8 yz + b_9 z^2 \\
\dot{z} &= c_0 + c_1 x + c_2 y + c_3 z + c_4 x^2 + c_5 xy + c_6 xz + c_7 y^2 + c_8 yz + c_9 z^2
\end{aligned}$$

- Up to 30 coefficients need to be identified using our training data.

Sparse Identification of Nonlinear Dynamics

Problem formulation: equivariant system

- System is equivariant w.r.t. to the transformation $(x, y, z) \mapsto (-x, -y, z)$. We thus have

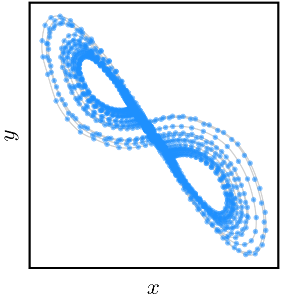
$$\gamma \cdot \dot{\boldsymbol{x}} = \boldsymbol{f}\left(\gamma \cdot \boldsymbol{x}\right)$$

with γ the matrix representation of the flip operator.

- The yet-unknown model reduces to

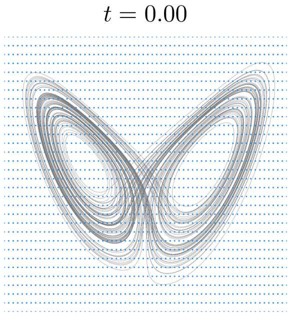
$$\begin{aligned}
\dot{x} &= a_0 x + a_1 y + a_2 xz + a_3 yz \\
\dot{y} &= b_0 x + b_1 y + b_2 xz + b_3 yz \\
\dot{z} &= c_0 + c_1 z + c_2 x^2 + c_3 xy + c_4 y^2 + c_5 z^2
\end{aligned}$$

- Only 14 terms need to be actually identified.



Sparse Identification of Nonlinear Dynamics

Problem formulation: dissipative system



- The system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$ is <u>dissipative</u>, thus

$$\nabla \cdot \boldsymbol{f}(\boldsymbol{x}) < 0 \quad \forall \boldsymbol{x}.$$

- It yields to the following constraints

$$a_0 + b_1 + c_1 < 0,$$
$$a_2 + b_3 + 2c_5 = 0.$$

- All three equations are coupled by these constraints and need to be identified jointly.

Sparse Identification of Nonlinear Dynamics

Problem formulation: energy-preserving quadratic nonlinearities

- Quadratic nonlinearities in Navier-Stokes are energy-preserving.
- For our system, energy equation would read

$$\dot{E} = c_0 z + a_0 x^2 + b_1 y^2 + c_1 z^2 + (a_1 + b_0) xy.$$

- The following constraints need to be satistifed

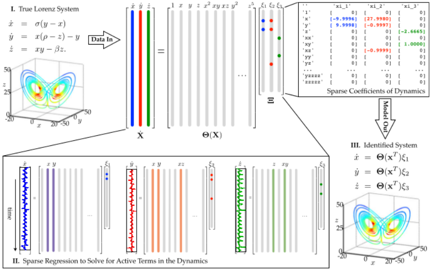
$$\begin{aligned} &b_3 + c_4 = 0, \quad a_2 + c_2 = 0 \\ &a_3 + b_2 + c_3 = 0. \end{aligned}$$

$$\begin{aligned} \text{State equation} &: \dot{\boldsymbol{x}} = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{x} + \mathcal{Q}(\boldsymbol{x}) \\ \text{Energy} &: E = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{x} \\ \text{Energy equation} &: \dot{E} = \boldsymbol{x}^T(\boldsymbol{b} + \boldsymbol{A}\boldsymbol{x}) \end{aligned}$$

Our optimization problem finally takes the following form

$$
\begin{aligned}
\underset{\boldsymbol{\xi}}{\text{minimize}} \; & \text{card}(\boldsymbol{\xi}) \\
\text{subject to} \; & \|\boldsymbol{\Theta}(\boldsymbol{x})\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2 < \sigma \\
& \boldsymbol{C}\boldsymbol{\xi} = 0 \\
& \boldsymbol{D}\boldsymbol{\xi} < 0.
\end{aligned}
$$

with $\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_x & \boldsymbol{\xi}_y & \boldsymbol{\xi}_z \end{bmatrix}^T$ the vector of unknown coefficients.



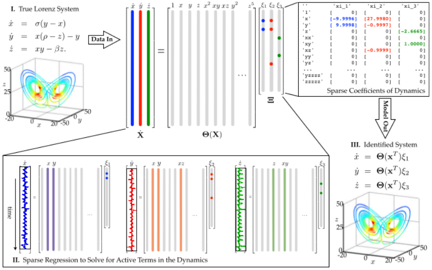
Sparse Identification of Nonlinear Dynamics

Final problem

In practice, this combinatorial problem is approximated by a convex relaxation, e.g.

$$
\begin{aligned}
\underset{\boldsymbol{\xi}}{\text{minimize}} \ & \|\boldsymbol{\Theta}(\boldsymbol{x})\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2 + \lambda R(\boldsymbol{\xi}) \\
\text{subject to } & \boldsymbol{C}\boldsymbol{\xi} = 0 \\
& \boldsymbol{D}\boldsymbol{\xi} < 0,
\end{aligned}
$$

with $R(\boldsymbol{x})$ some sparsity promoting heuristic. Here, we chose $R(\boldsymbol{x}) = \|\boldsymbol{\xi}\|_2^2$ (i.e. Thikonov regularization).



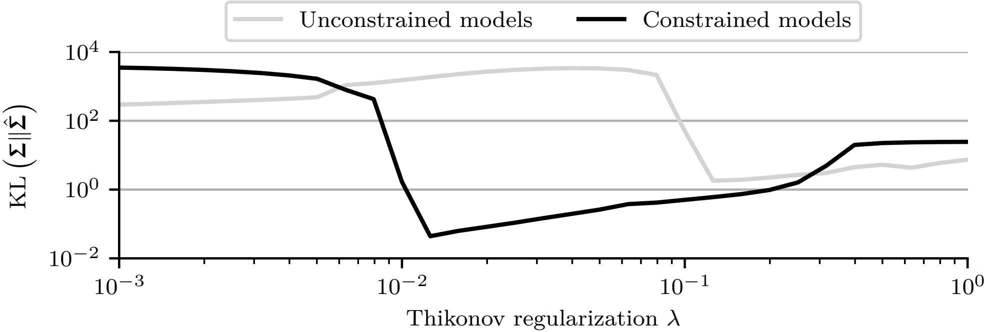
Sparse Identification of Nonlinear Dynamics

Identifying the model

- Our dataset spans 500τ time-units. A sequence of 400τ is used for training while the remaining 100τ are used for testing.
- The optimization problem is implemented using the python bindings of `cvxopt`.
- Model selection is based on the ability of the identified system to reproduce the second-order statistics of the true system.

Sparse Identification of Nonlinear Dynamics

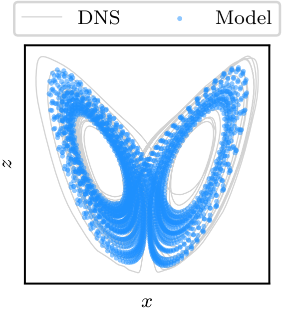
Identified systems and model selection



$$\mathrm{KL}(\boldsymbol{A}|\boldsymbol{B}) = \frac{1}{2}\left(\mathrm{Tr}(\boldsymbol{B}^{-1}\boldsymbol{A}) - n + \ln\frac{|\boldsymbol{B}|}{|\boldsymbol{A}|}\right)$$

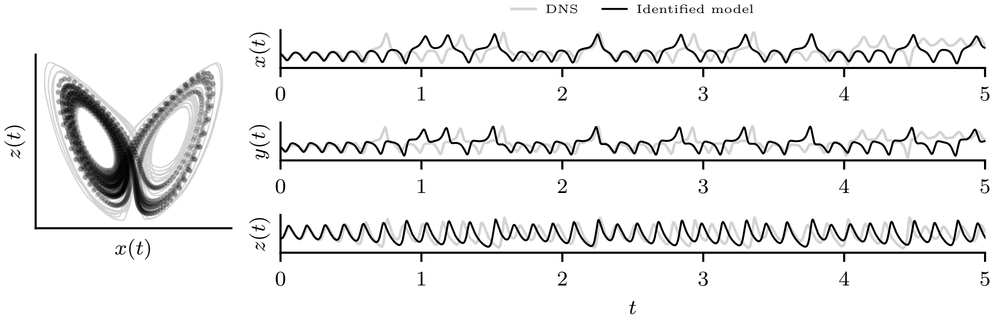
Selected model

$$\dot{x} = -76.08x + 88.39y$$
$$\dot{y} = 20.76x - 4.19y - 41.49xz$$
$$\dot{z} = -43.67 - 17.31z + 41.49xy.$$



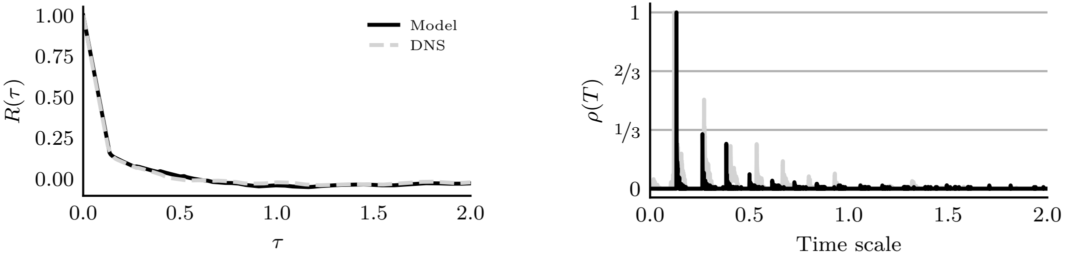
- <u>Lorenz-like</u> system satisfying all of the physical constraints by design.
- Every single term can be intepreted and associated to a particular physical mechanism.

Selected model



Sparse Identification of Nonlinear Dynamics

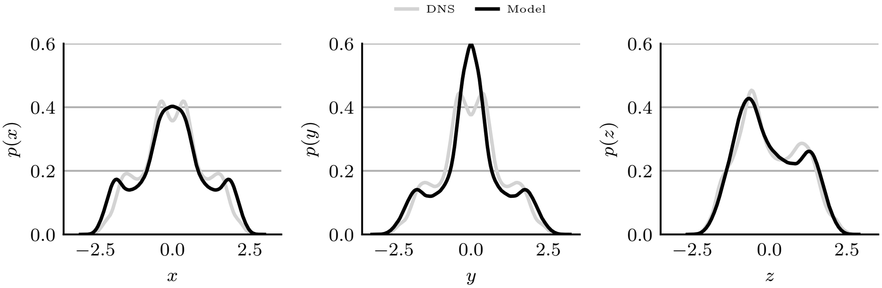
Comparing the statistical properties



The model accurately captures the dynamical properties of the switches.

Sparse Identification of Nonlinear Dynamics

Comparing the statistical properties

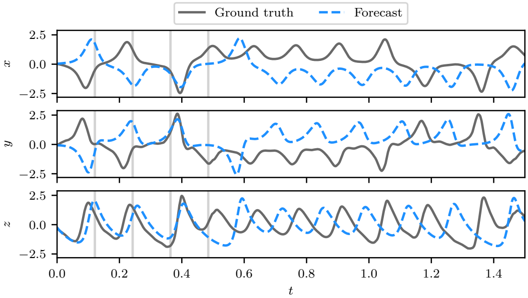


Overall agreements regarding the probability density functions albeit there is a small (yet unexplained) mismatch in the center.

Sparse Identification of Nonlinear Dynamics

Forecasting abilities

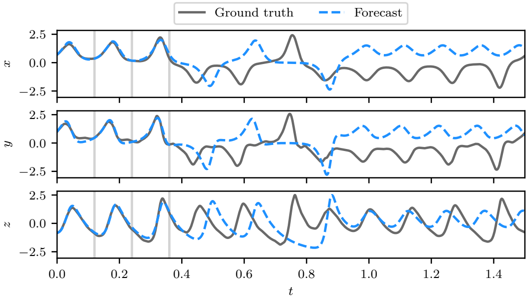
- The identified model can easily be used for forecasting.
- Some specific regions of phase space are prone to large forecasting errors.
- Forecasts are statistically accurate on the characteristic time-scale τ.



Sparse Identification of Nonlinear Dynamics

Forecasting abilities

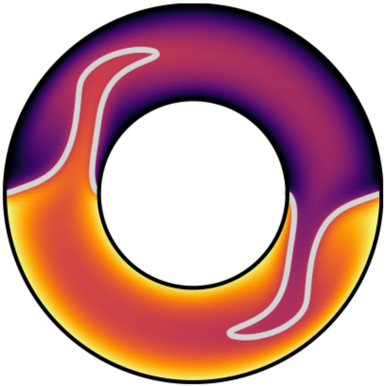
- The identified model can easily be used for forecasting.
- Some specific regions of phase space are prone to large forecasting errors.
- Forecasts are statistically accurate on the characteristic time-scale τ.



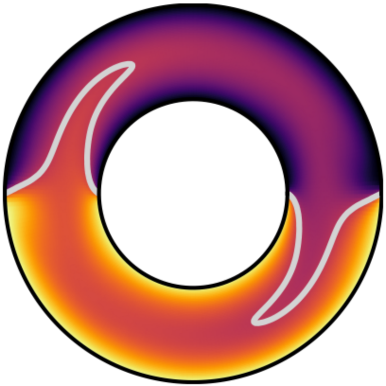
Sparse Identification of Nonlinear Dynamics

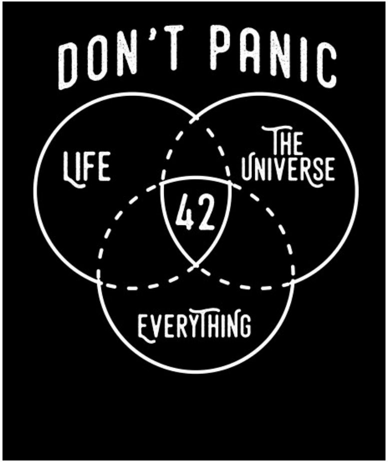
Eyeball-norm comparison

Direct numerical simulation



Reduced-order model



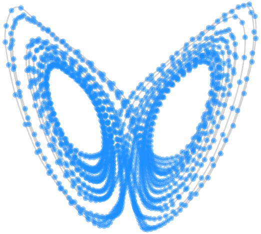


Part IV

Random thoughts on life, the universe and everything

- Dimensionality reduction as a preprocessing step to model the dynamics of physical systems has two competing objectives
    1. Project the data in a low-dimensional latent space.
    2. The resulting coordinate system needs to be well-suited for modeling dynamics.

- While goal n°1 can easily be formalized mathematically, goal n°2 is more abstract.



Random thought on life, the universe and everything

Dimensionality reduction

- Given full-state data, personnal experience suggests that DMD-like techniques are well-suited.
  - ↪ Relies on highly-efficient numerical linear algebra techniques.
  - ↪ Inherently interpretable low-dimensional phase space.

- From a dynamical system point of view, DMD has been related to the Koopman operator.
  - ↪ Provides a coordinate system where the dynamics are approximately linear.

Learning the input-output map $\neq$ Learning the physics

?

- Even on this "simple" example, basic ML algorithms are unable to learn the physical properties of the system.
- Explicitely enforcing prior physical knowledge leads to easier training and orders of magnitude more accurate models.

How to scale things up ?

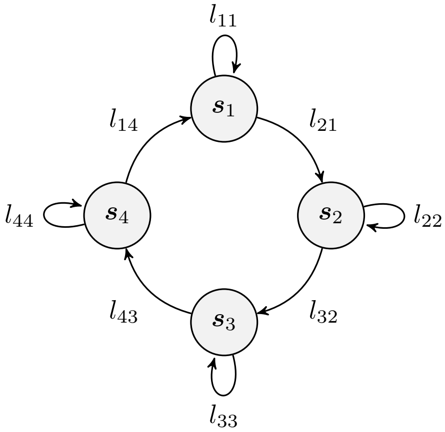
↪ Equivariant autoencoders to encode existing spatial symmetries in the dimensionality reduction.

↪ Treating a neural network as a function $\boldsymbol{f}(\boldsymbol{x})$, use ∂ifferential programming to apply differential operators (e.g. $\nabla \cdot \boldsymbol{f}(\boldsymbol{x})$, etc).

↪ Encode prior physical knowledge in the architecture / loss of the neural networks (see recent *Lagrangian neural networks* for instance).

Thank you for your attention.

Any questions?



Appendix

Miscellaneous

- Dynamic Mode Decomposition:
  - ↪ Theorem & Proof
  - ↪ PCA vs. CCA vs. DMD
- Probabilistic model:
  - ↪ How to choose the number of clusters ?
- SINDy with constraints:
  - ↪ Covariance matrix of the parameters.

Dynamic Mode Decomposition

Theorem : Optimal solution to the DMD problem

The solution to the DMD problem

$$
\begin{aligned}
\underset{\boldsymbol{P},\boldsymbol{Q}}{\text{minimize}} \ & \|\boldsymbol{Y} - \boldsymbol{P}\boldsymbol{Q}^H\boldsymbol{X}\|_F^2 \\
\text{subject to } & \mathsf{rank}\ \boldsymbol{P} = \mathsf{rank}\ \boldsymbol{Q} = r \\
& \boldsymbol{P}^H\boldsymbol{P} = \boldsymbol{I}
\end{aligned}
$$

is given by the first r eigenvectors of the symmetric positive definite matrix $\boldsymbol{C}_{\boldsymbol{yx}}\boldsymbol{C}_{\boldsymbol{xx}}^{-1}\boldsymbol{C}_{\boldsymbol{xy}}$.

Dynamic Mode Decomposition

Thereom & Proof

Proof : Haven't had time to write it properly in beamer !

Principal Component Analysis

$$\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} \Lambda = \begin{bmatrix} C_{yy} & 0 \\ 0 & C_{xx} \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix}$$

Canonical Correlation Analysis

$$\begin{bmatrix} C_{yy} & 0 \\ 0 & C_{xx} \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} \Lambda = \begin{bmatrix} 0 & C_{yx} \\ C_{xy} & 0 \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix}$$

Dynamic Mode Decomposition

$$\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix} \Lambda = \begin{bmatrix} 0 & C_{yx} \\ C_{xy} & -C_{xx} \end{bmatrix} \begin{bmatrix} P \\ Q \end{bmatrix}$$

Probabilistic model

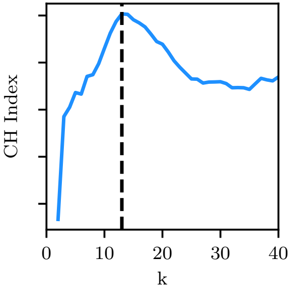
How to choose the number of clusters ?

- Various heuristics exist to assess how good our clustering is.
- Calinski & Harabasz Index : choose the number of clusters based on the ratio of within-cluster and between-cluster dispersions

$$CH(k) = \frac{tr(\boldsymbol{B}_k)}{tr(\boldsymbol{W}_k)} \times \frac{n_E - k}{k - 1}$$

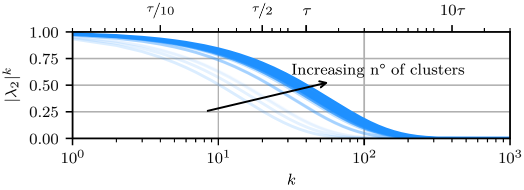
  where $tr(\boldsymbol{B}_k)$ is the between-cluster dispersion and $tr(\boldsymbol{W}_k)$ is the within-cluster dispersion.

- Other possible metrics : Silouhette coefficient, Davies-Bouldin Index, etc.



Sparse Identification of Nonlinear Dynamics

Covariance matrix of the parameters



- Dynamics of the system are correlated only on a time-scale $\mathcal{O}(\tau)$.
- Dataset can be partionned into n independant realizations.

- One can use bootstrapping to estimate the covariance matrix of the parameters.
- It then enables us to perform uncertainty quantification when forecasting.