Learning Koopman Invariant Subspaces for Dynamic Mode Decomposition

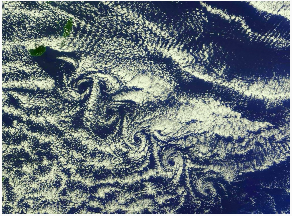
Naoya Takeishi (The University of Tokyo),
Yoshinobu Kawahara (Osaka University / RIKEN AIP),
Takehisa Yairi (The University of Tokyo)

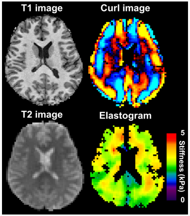
IBIS 2018 in Tokyo, November 2017

To appear in NIPS 2017.
preprint: https://arxiv.org/abs/1710.04340

Background

Motivation: Analysis of Dynamical Systems



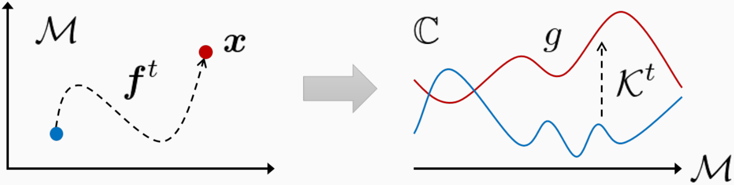


- In physics, biology, etc., a wide variety of complex phenomena are described in terms of **dynamical systems**.
- In machine learning, **state space models** are utilized classically.

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t), \quad \boldsymbol{x} \in \mathcal{M} \text{ (state space)}$$

☹ Challenging to know (a part of) properties of nonlinear $\boldsymbol{f}$.

Background: Operator-theoretic view of nonlinear dynamics



- Fundamental idea: Lift nonlinear system to linear regime by $\mathcal{K}$.

Definition (Koopman operator [Koopman '31, Mezić '05])

Let $g : \mathcal{M} \to \mathbb{C}$ be an observable. **Koopman operator** $\mathcal{K}$ is an infinite dimensional linear operator that describes time-evolution of g, i.e.,

$$g(\boldsymbol{f}(\boldsymbol{x})) = \mathcal{K}g(\boldsymbol{x}), \quad g \in \mathcal{G} \text{ (function space)}$$

☺ Benefit: control, **modal decomposition** ($\to$ next page), etc.

Modal decomposition based on spectra of $\mathcal{K}$

- Consider eigenvalues λ and eigenfunctions φ of $\mathcal{K}$:

$$\mathcal{K}\,\varphi_i(\boldsymbol{x}) = \lambda_i\,\varphi_i(\boldsymbol{x}) \quad \text{for} \quad i = 1, 2, \ldots, \infty.$$

- For simplicity, assume $\mathcal{K}$ has only discrete spectrum (eigenvalues).
- Project $g(\cdot)$ to $\mathrm{span}\{\varphi_1(\cdot), \varphi_2(\cdot), \ldots\} \to$ Koopman modes

$$g(\boldsymbol{x}) = \sum_{i=1}^{\infty} \varphi_i(\boldsymbol{x})\, v_i$$

- Applying $\mathcal{K}$ to both sides for t times, we have $g(\boldsymbol{x}_t)$**'s decomposition into quasi-periodic modes** (Koopman mode decomposition, KMD):

$$g(\boldsymbol{x}_t) = \sum_{i=1}^{\infty} \lambda_i^t \underbrace{\varphi_i(\boldsymbol{x}_0) v_i}_{w_i}, \quad \text{where} \quad \begin{cases} |\lambda_i| & \text{decay rate of } w_i, \\ \angle\lambda_i & \text{frequency of } w_i. \end{cases}$$

Application examples

- fluid mechanics [many], robotics [Berger+ ’15], neuroscience [Brunton+ ’16], epidemiology [Proctor&Eckhoff ’15], finance [Mann&Kutz ’16], medical care [Bourantas+ ’16], building systems [Georgescu+ ’12], power systems [Raak+ ’16, Susuki+ ’16], image processing [Kutz+ ’16], control [Mauroy&Goncalves ’16], etc.

How to compute modal decomposition by Koopman operator?

- KMD can be computed using **dynamic mode decomposition (DMD)** [Rowley+ '09, Schmid 10, etc.] if the assumption below is satisfied.
  - DMD just computes eigendecomposition of $\boldsymbol{A}$, where $\boldsymbol{y}_{t+1} \approx \boldsymbol{A}\boldsymbol{y}_t$.
  - $\boldsymbol{A}$'s eigvalues and eigvectors converge to $\mathcal{K}$'s eigvalues and modes.

Assumption (Dataset from $\mathcal{K}$-invariant subspace [Budišić+ '12])

Time-series data $(\boldsymbol{y}_0, \dots, \boldsymbol{y}_m)$ are generated by

$$\forall t,\ \boldsymbol{y}_t = \boldsymbol{g}(\boldsymbol{x}_t) \quad \text{where} \quad \boldsymbol{g} = [g_1 \ \cdots \ g_n]^\mathsf{T},$$

and $\{g_1, \dots, g_n\}$ spans a subspace that is invariant to $\mathcal{K}$, i.e.,

$$\exists G \subset \mathcal{G} \ \text{ s.t. } \forall g \in G,\ \mathcal{K}g \in G \quad \text{and} \quad \operatorname{span}\{g_1, \dots, g_n\} = G.$$

☹ Of course, this assumption is not satisfied generally.

Learning Koopman Invariant Subspaces from Data

To compute KMD, we need data from $\mathcal{K}$-invariant subspace!

- "Data from $\mathcal{K}$-invariant subspace" is not trivial ☹
- Previous approaches:
  - forget it when data is high-dimensional (e.g. CFD)
  - transform data by nonlinear basis functions [Williams+ '15]
  - define KMD for observables in RKHS [Kawahara '16]

- Our approach: Learn a $\mathcal{K}$-invariant subspace using data.

Solution: Idea

Theorem ($\mathcal{K}$-invariant subspace)

A set of observables $\{g_1, \ldots, g_n\}$ spans a $\mathcal{K}$-invariant subspace if and only if $\boldsymbol{g} = [g_1 \cdots g_n]^\top$ and $\boldsymbol{g} \circ \boldsymbol{f}$ are linearly dependent.

- Main issue: Learn $\boldsymbol{g}$ that makes $\boldsymbol{g}$ and $\boldsymbol{g} \circ \boldsymbol{f}$ linearly dependent.
  $\rightarrow$ Minimize **residual sum of squares of linear LS model**:

$$
\begin{aligned}
\mathcal{L}_{\mathsf{RSS}}(\boldsymbol{g}; \boldsymbol{x}_{0:m}) &= \|\boldsymbol{Y}_1 - (\boldsymbol{Y}_1 \boldsymbol{Y}_0^\dagger)\boldsymbol{Y}_0\|_{\mathrm{F}}^2, \\
\boldsymbol{Y}_0 &= [\boldsymbol{g}(\boldsymbol{x}_0) \ \cdots \ \boldsymbol{g}(\boldsymbol{x}_{m-1})], \\
\boldsymbol{Y}_1 &= [\boldsymbol{g}(\boldsymbol{x}_1) \ \cdots \ \boldsymbol{g}(\boldsymbol{x}_m)]
\end{aligned}
$$

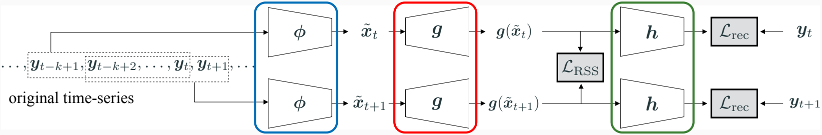
- And minor modifications to:
  - Estimate $\boldsymbol{x}_t$ from $\boldsymbol{y}_{t-k+1:t}$ (delay-coordinate embedding [Takens '81]).
    - because $\boldsymbol{y}$ is often rank-deficient
  - Prevent trivial $\boldsymbol{g}$ by adding a term to recover $\boldsymbol{y}$ from the values of $\boldsymbol{g}$.

Solution: Implementation

- The loss function for learning $\mathcal{K}$-invariant subspace:

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{g}, \boldsymbol{h};\ \boldsymbol{y}) = \tilde{\mathcal{L}}_{\mathsf{RSS}}(\boldsymbol{\phi}, \boldsymbol{g};\ \boldsymbol{y}) + \alpha\ \underbrace{\mathcal{L}_{\mathsf{rec}}(\boldsymbol{g}, \boldsymbol{h};\ \boldsymbol{y})}_{\sum_t \|\boldsymbol{h}(\boldsymbol{g}_t) - \boldsymbol{y}_t\|_2^2}$$

- Modeling with multilayer perceptrons for each of
  - $\boldsymbol{\phi}$: estimates $\boldsymbol{x}_t$ from $\boldsymbol{y}_{t-k+1:t}$ (delay-coordinate embedder)
  - $\boldsymbol{g}$:spans $\mathcal{K}$-invariant subspace
  - $\boldsymbol{h}$: recovers $\boldsymbol{y}$ from values of $\boldsymbol{g}$ (to prevent trivial solution)



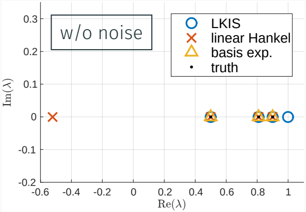
Experiment example

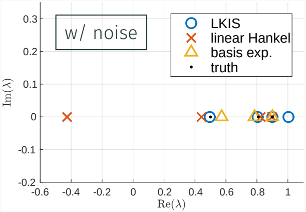
- Generate data from a toy system ($\lambda = 0.9,\ \mu = 0.5$):

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}) = \begin{cases} \lambda x_{1,t}, \\ \mu x_{2,t} + (\lambda^2 - \mu) x_{1,t}^2 \end{cases}$$

  - $\mathcal{K}$-invariant subspace of this system is $\mathrm{span}\{x_1, x_2, x_1^2\}$, and the corresponding eigenvalues are λ, μ, λ^2 (and $\lambda^i \mu^j$).

- Want to estimate the Koopman eigenvalues;
  - Proposed method (LKIS) successes even with noise in data.
  - Nonlinear basis function (using $\{x_1^2\}$) fails with noise.

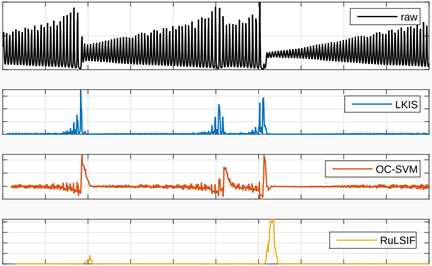




Application example: unstable phenomena detection

- The Koopman eigenfunction with small $|\lambda|$ corresponds to a **rapidly-decaying component**.
  - Watching such eigenfunction, we can detect unstable phenomena.

- Example: laser pulsation data [Santa Fe Time-Series Dataset A]
  - Plotting learned eigfun. with small $|\lambda|$, we see the peaks are corresponding to the rapid change of amplitude.
  - Change-point detection algorithms output similar results.



Conclusion

Summary

Koopman analysis

- lifts analysis of nonlinear dynamical systems to linear regime by defining Koopman operator.
- Benefit: modal decomposition, control, etc.

Dynamic mode decomposition

- can approximate KMD if data is generated with observables that span $\mathcal{K}$-inv. subspace.

In this work, we learn $\mathcal{K}$-inv. subspace from data

- RSS minimization with neural networks

Will be presenting at Poster D1-1 tomorrow!

preprint: `https://arxiv.org/abs/1710.04340`

implementation: `https://github.com/thetak11/learning-kis`