The Koopman operator in system identification

Some quantitative remarks

Christophe Zhang

Chair for Applied Analysis, Alexander von Humboldt professorship

Mini-workshop on hyperbolic problems, Oct.12 2020

Summary

Systems identification

Systems identification

$$\begin{cases} \dfrac{\partial}{\partial t} u + u \cdot \quad u - \nu \Delta u = - \quad p, \\ \quad \cdot u = 0. \end{cases}$$

Systems identification

$$\begin{cases} \dfrac{\partial}{\partial t} u + u \cdot \nabla u - \nu \Delta u = -\nabla p, \\ \nabla \cdot u = 0. \end{cases}$$

Goal

Given a system for which the dynamics are unknown, **learn** a model from data, to make **predictions** for the evolution of the system.

The Koopman operator

Introduced by B.O. Koopman in *Hamiltonian systems and transformation in Hilbert space*. Proc. Natl. Acad. Sci. U. S. A.17(5), 315 (1931)

A nonlinear vector field $V : \mathrm{R}^d \to \mathrm{R}^d$ can define a nonlinear ODE

$$\dot{x} = V(x)$$

as well as the *linear* transport equation

$$\phi_t = V(x) \cdot \quad \phi.$$

The **Koopman operator** is the unbounded operator

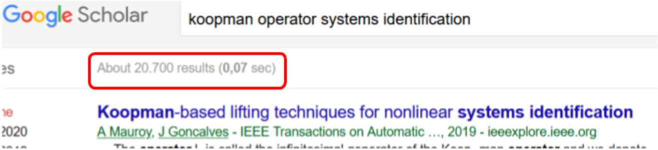
$$K := V(x) \cdot \quad .$$

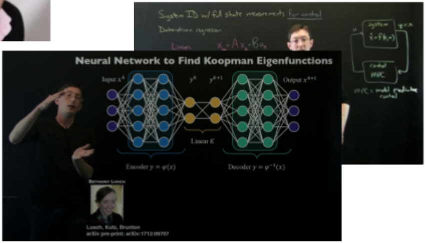
Semigroup formulation: note Φ_V^t the flow of V.

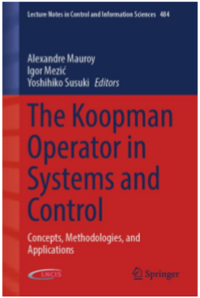
Then the **Koopman semigroup** is given by

$$U(t) : \phi \to \phi \circ \Phi_V^t.$$

The new life of the Koopman operator







Koopman operator for system identification

System identification: learn V from data.

Leverage the linearity of the Koopman operator:

$$\dot{x} = V(x) \leftrightarrow \phi_t = V(x) \cdot \quad \phi.$$

Koopman operator for system identification

System identification: learn V from data.

Leverage the linearity of the Koopman operator:

$$\dot{x} = V(x) \leftrightarrow \phi_t = V(x) \cdot \quad \phi.$$

Koopman operator for system identification

System identification: learn V from data.

Leverage the linearity of the Koopman operator:

$$\dot{x} = V(x) \leftrightarrow \phi_t = V(x) \cdot \quad \phi.$$

We can't get the "whole" transport equation!

Koopman operator for system identification

System identification: learn V from data.

Leverage the linearity of the Koopman operator:

$$\dot{x} = V(x) \leftrightarrow \phi_t = V(x) \cdot \quad \phi.$$

We can't get the "whole" transport equation!

Idea

Learn a *Galerkin approximation* of the Koopman operator K from data.

Koopman operator for system identification

System identification: learn V from data.

Leverage the linearity of the Koopman operator:

$$\dot{x} = V(x) \leftrightarrow \phi_t = V(x) \cdot \quad \phi.$$

We can't get the "whole" transport equation!

Idea

Learn a *Galerkin approximation* of the Koopman operator K from data.

Learning nonlinear dynamics

.

Koopman operator for system identification

System identification: learn V from data.

Leverage the linearity of the Koopman operator:

$$\dot{x} = V(x) \leftrightarrow \phi_t = V(x) \cdot \quad \phi.$$

We can't get the "whole" transport equation!

Idea

Learn a *Galerkin approximation* of the Koopman operator K from data.

~~Learning nonlinear dynamics~~

Learning a linear, finite-dimensional reduction of the associated linear transport equation.

Summary

Take a subspace $H_N \quad L^2$, $\dim(H_N) = N$. Then the ODE

$$\phi_t = \Pi_N K\phi, \quad \phi(0) \quad H_N,$$

is the Galerkin approximation of the transport equation

$$\phi_t = V(x) \cdot \quad \phi$$

on the subspace H_N.

Numerical analysis	Koopman method
Galerkin approximation	Trajectory data of an unknown ODE
↓	↓
Simulate a known PDE	Identify Galerkin approximation

Linear regression

For a given basis $\varphi_1, \cdots, \varphi_N$ of H_N, the Galerkin approximation writes in matrix form

$$\dot{X} = K_N X, \quad X \in \mathbb{R}^N.$$

Finding K_N is a simple linear regression: we need to know how the trajectories starting from the basis vectors (i.e. the φ_j) evolve in time.

Linear regression

For a given basis $_1; \quad ; _N$ of H_N, the Galerkin approximation writes in matrix form

$$\dot{X} = K_N X; \quad X \in \mathbb{R}^N.$$

Finding K_N is a simple linear regression: we need to know how the trajectories starting from the basis vectors (i.e. the $_j$) evolve in time.

Not so simple

We do not have this information, we have to sample it from data.

generator Extended Dynamic Mode Decomposition

Introduced in Williams, M.O., Kevrekidis, I.G., Rowley, C.W. A Data Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. J Nonlinear Sci 25, 1307 1346 (2015).

gEDMD

Pick m random points $(x_n)_{1 \le n \le m}$, data set $\{ x_n ; x_n(0) ; \; 1 \le n \le m \}$

gEDMD

Pick m random points $(x_n)_{1\le n\le m}$, data set $\{x_n; V(x_n); 1 \le n \le m\}$

gEDMD

Pick m random points $(x_n)_{1 \le n \le m}$, data set $\{x_n; V(x_n); 1 \le n \le m\}$

$$\Psi(X) := \begin{pmatrix} \psi_1(x_1) & \cdots & \psi_1(x_m) \\ \vdots & \ddots & \vdots \\ \psi_N(x_1) & \cdots & \psi_N(x_m) \end{pmatrix};$$

$$d\Psi(Y) := \begin{pmatrix} x_{\underline{1}} \cdot \nabla \psi_1(x_1) & \cdots & x_{\underline{m}} \cdot \nabla \psi_1(x_m) \\ \vdots & \ddots & \vdots \\ x_{\underline{1}} \cdot \nabla \psi_N(x_1) & \cdots & x_{\underline{m}} \cdot \nabla \psi_N(x_m) \end{pmatrix}$$

Proposition (Williams, Kevrekidis, Rowley, 2015)

The solution A_N^m of the least squares problem:

$$A_N^m := \operatorname{argmin}_{A \in M_N(\mathbb{R})} \| A\ (\ X) \qquad (\ Y) \|^2$$

satis es

$$A_N^m \xrightarrow[m \to \infty]{} \quad K_N :$$

For a large enoughm (Monte-Carlo convergence rate...), solving a least squares problem gives us a goodapproximation of the reduced-order model of the Koopman transport equation.

Another formulation

We want K_N the matrix of $_N K \in L(V_N)$.

Matrix representation of the projection:

$$' = \sum_{i=1}^{N} c_i \ _i = c^{\top} \ ; \qquad _N K' = \sum_{i=1}^{N} c_i \ _N K \ _i = c^{\top} K_N \ :$$

$$K_N = \begin{pmatrix} \langle K \ _1; \ _1 \rangle & & \langle K \ _1; \ _N \rangle \\ \vdots & \ddots & \vdots \\ \langle K \ _N; \ _1 \rangle & & \langle K \ _N; \ _N \rangle \end{pmatrix} \begin{pmatrix} \langle \ _1; \ _1 \rangle & & \langle \ _1; \ _N \rangle \\ \vdots & \ddots & \vdots \\ \langle \ _N; \ _1 \rangle & & \langle \ _N; \ _N \rangle \end{pmatrix}^{-1}$$

Approximate the integrals of the scalar products $\to$ approximate K_N.

Yields an operator on H_N:

$$K_N^m \quad _N K$$

How good is this data-driven approximation?

$$\|K\varphi - K_N^m\varphi\|_{L^2} \leq \underbrace{\|(\Pi_N - I)K\varphi\|_{L^2}}_{} + \underbrace{\|(\Pi_N K - K_N^m)\varphi\|_{L^2}}_{\text{Monte-Carlo method}}$$

For trajectories: initial condition $\varphi_0 \in H_N$

$$\|\varphi(t;\cdot) - \varphi_N^m(t;\cdot)\|_{L^2} \leq \underbrace{\|\varphi(t;\cdot) - \varphi_N(t;\cdot)\|_{L^2}}_{} + \underbrace{\|\varphi_N(t;\cdot) - \varphi_N^m(t;\cdot)\|_{L^2}}_{\text{Monte-Carlo method}}$$

How good is this data-driven approximation?

$$\|K\varphi - K_N^m\varphi\|_{L^2} \le \underbrace{\|(\Pi_N - I)K\varphi\|_{L^2}}_{\text{Quality of Galerkin approx}} + \underbrace{\|(\Pi_N K - K_N^m)\varphi\|_{L^2}}_{\text{Monte-Carlo method}}$$

For trajectories: initial condition $\varphi_0 \in H_N$

$$\|\varphi(t;\cdot) - \varphi_N^m(t;\cdot)\|_{L^2} \le \underbrace{\|\varphi(t;\cdot) - \varphi_N(t;\cdot)\|_{L^2}}_{\text{Quality of Galerkin approx}} + \underbrace{\|\varphi_N(t;\cdot) - \varphi_N^m(t;\cdot)\|_{L^2}}_{\text{Monte-Carlo method}}$$

Choosing the Galerkin approximation

We can't get the whole transport equation!

Choosing the Galerkin approximation

We don't need the whole transport equation!

$$\dot{x} = V(x): \tag{1}$$

We don't need the whole transport equation!

$$\underline{x} = V(x): \qquad (1)$$

Solutions of the transport equation:

$$' (t; \;) = '_0 \quad {}^{t}_{V}:$$

Then for $'_0 = e_i$,

$$' (t; x) = \quad {}^{t}_{V}(x) \;_i; \quad K e_i = V \;\; r \; e_i = V_i:$$

We only need these initial conditions!

Choosing the Galerkin approximation

We don't need the whole transport equation!

$$\underline{x} = V(x): \tag{1}$$

Solutions of the transport equation:

$$'\,(t;\ ) = '_0 \quad {}^{t}_{V}:$$

Then for $'_0 = e_i$,

$$'\,(t;x) = \quad {}^{t}_{V}(x)_{\ i}; \quad Ke_i = V \quad r\ e_i = V_i:$$

We only need these initial conditions! A good choice of Galerkin approximation:

- H_N contains the coordinate functions.
- The Galerkin approximation of the transport equation is accurate for these initial conditions.

Ideal case:

$$_N K\varphi = K\varphi, \quad \forall \varphi \in H_N.$$

The Galerkin approximation is exact for any initial condition in H_N.

Good Galerkin approximations?

Ideal case:

$$\Pi_N K\varphi = K\varphi, \quad \forall \varphi \in H_N.$$

The Galerkin approximation is exact for any initial condition in H_N.

An example that keeps popping up:

$$\begin{aligned} \dot{x}_1 &= \mu x_1, \\ \dot{x}_2 &= \lambda(x_2 - x_1^3). \end{aligned} \qquad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_1^3 \end{pmatrix}$$

then,

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & 3\mu \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}. \tag{2}$$

See Brunton SL, Brunton BW, Proctor JL, Kutz JN. Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control. PLoS One. 2016 Feb 26;11(2):e0150171.

Good Galerkin approximations?

Ideal case:

$$\Pi_N K\varphi = K\varphi, \quad \forall \varphi \in H_N.$$

The Galerkin approximation is exact for any initial condition in H_N.
An example that keeps popping up:

$$\dot{x} = P(x), \quad x \in \mathbb{R}$$

then,

$$\deg(K^n x) = n(\deg(p) - 1) + 1,$$

See Brunton SL, Brunton BW, Proctor JL, Kutz JN. Koopman Invariant Subspaces and Finite Linear Representations of Nonlinear Dynamical Systems for Control. PLoS One. 2016 Feb 26;11(2):e0150171.

One size ts all:

Without any prior knowledge on V (except regularity), it seems more reasonable to look at classical approximation spaces.

Linear nite elements:

$$k(\ \ _N \quad I)K'\ k_{L^2} \quad Ch^2 kr\ (K'\ )k_{L^2};$$

$$k'\ (t;\ ) \quad '\ _N(t;\ )k_{L^2} \quad Ch \quad k'\ k_{L^2} + \int_0^t kK'\ (s;\ )k_{L^2}ds$$

Classical approximation spaces

One size ts all:

Without any prior knowledge on V (except regularity), it seems more reasonable to look at classical approximation spaces.

Linear nite elements:

$$k(\ _N \quad I)K'\ k_{L^2} \quad CN^{\ \frac{2}{d}} kr\ (K'\ )k_{L^2};$$

$$k'\ (t;\ ) \quad '_N(t;\ )k_{L^2} \quad CN^{\ \frac{1}{d}}\ k'\ k_{L^2} + \int_0^t kK'\ (s;\ )k_{L^2} ds$$

One size ts all:

Without any prior knowledge on V (except regularity), it seems more reasonable to look at classical approximation spaces.

Linear nite elements:

$$\|(\ _N \quad I)K\varphi\|_{L^2} \quad CN^{\ \frac{2}{d}}\|\nabla\ (K\varphi\ )\|_{L^2};$$

$$\|\varphi\ (t;\ ) \quad \varphi_N(t;\ )\|_{L^2} \quad CN^{\ \frac{1}{d}}\ \ \|\varphi\ \|_{L^2} + \int_0^t \|K\varphi\ (s;\ )\|_{L^2}\,ds$$

Generality comes at a cost   : here, the curse of dimensionality.

In general, sub-optimality.

see Kurdila, A. J. and Bobade, P., *Koopman Theory and Linear Approximation Spaces*, arXiv e-print, 2018.

To sum up

Summary

Is the promise held?

We implemented gEDMD on ODEs with simple nonlinearities.

We focus on the identification of the vector field $V(x)$ of the ODE.

We use two different subspaces for the Galerkin approximation of the Koopman transport equation:

Linear finite elements P_1 and **Chebyshev polynomials** T_n.

Compared with classical interpolation (linear and splines).

Parameters:

N is the dimension of the subspace of functions (number of nodes for P_1, max degree for T_n).

m is the number of samples. The ration m/N plays an important role.

Credits: Jesus Oroya, CCM, University of Deusto.

Finite elements

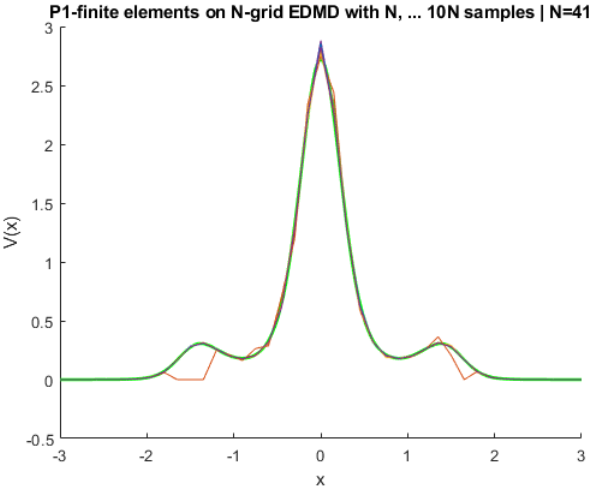
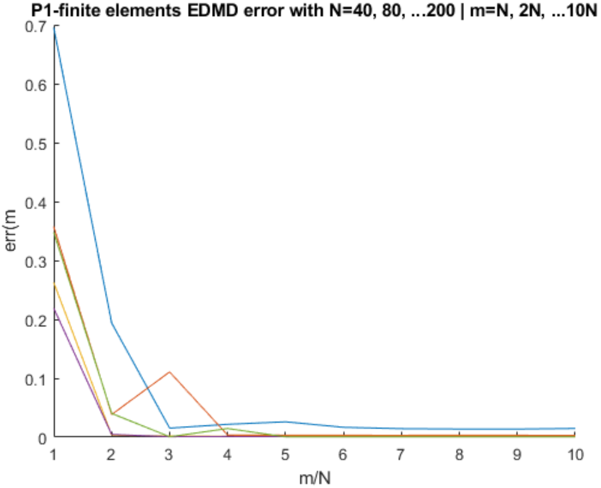


Figure 1: recovery of $V(x) = \exp(-x^2 + \cos(4x))$ on $[-3, 3]$.

Finite elements

How does the error decrease with the number of samples for given values of N?



Finite elements

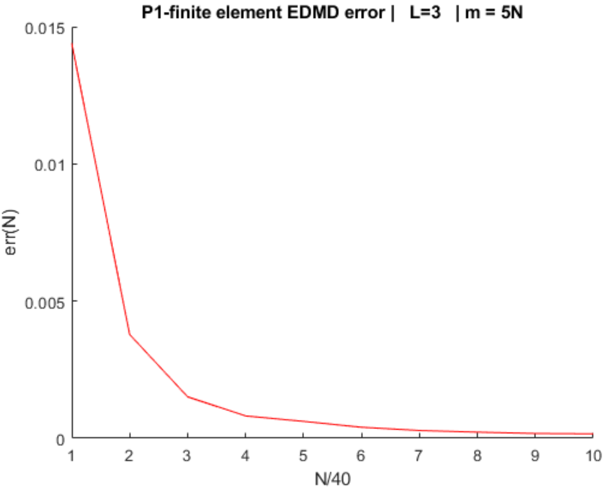


Figure 3: Recovery of $V(x) = \exp(-x^2 + \cos(4x))$ on $[-3, 3]$.

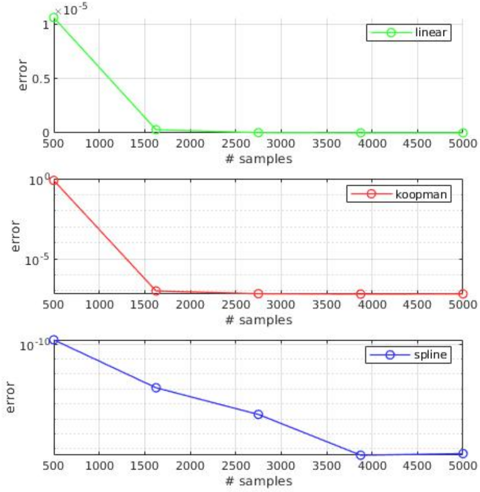


Figure 4: Comparison of three methods with $V(x) = \sin(x)$.

Chebyshev polynomials

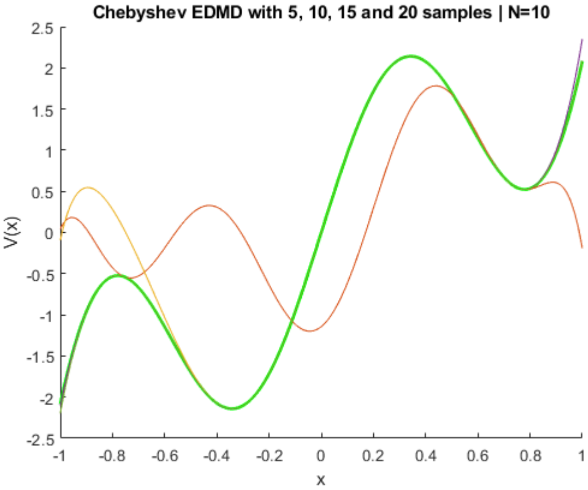


Figure 5: recovery of $V(x) = 4x^3 + 2\sin(5x)$ on $[-3, 3]$.

Chebyshev polynomials

Compare again with splines and linear interpolation:

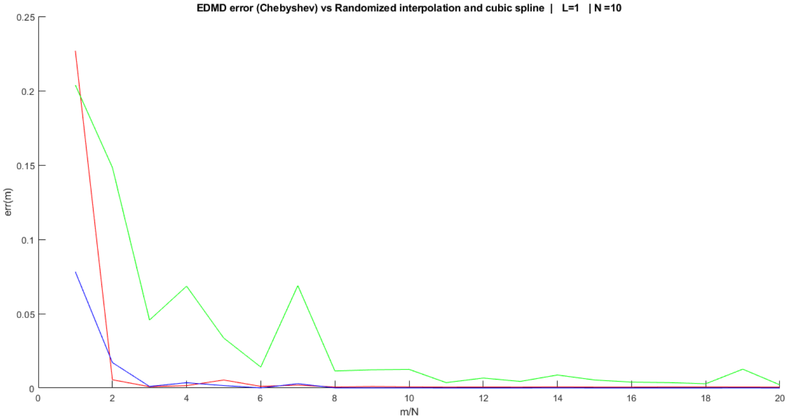


Figure 6: recovery of $V(x) = 4x^3 + 2\sin(5x)$ on $[-3, 3]$.

Chebyshev polynomials

Compare again with splines and linear interpolation: focus on high number of samples

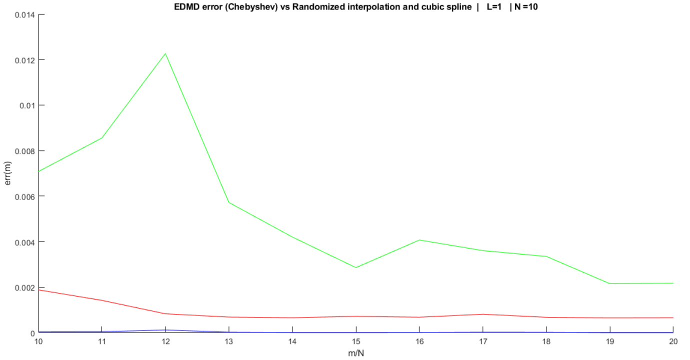


Figure 7: recovery of $V(x) = 4x^3 + 2\sin(5x)$ on $[-3, 3]$.

To conclude

Conclusion

Performance of EDMD with classical approximation spaces: probably not worth it! Unless...

Conclusion

Performance of EDMD with classical approximation spaces: probably not worth it! Unless...

An intermediary between the utopia (H_N invariant) and the bazooka ($H_N = \mathsf{P}_1, \cdots$).

Conclusion

Performance of EDMD with classical approximation spaces: probably not worth it! Unless...

An intermediary between the utopia (H_N invariant) and the bazooka ($H_N = \mathsf{P}_1, \cdots$).

The optimal ***relevant*** H_N of low dimension?

Conclusion

Performance of EDMD with classical approximation spaces: probably not worth it! Unless...

An intermediary between the utopia (H_N invariant) and the bazooka ($H_N = \mathsf{P}_1, \cdots$).

The optimal ***relevant*** H_N of low dimension?

Approximations of V and the trajectories cheap to compute.

Challenge:

- Characterize the existence of such spaces for a given system.
- Design data-driven methods to find them if/when they exist (already some deep-learning attempts...).

Thank you!

Vielen Dank!

Some references

EDMD: Williams, M.O., Kevrekidis, I.G., Rowley, C.W. *A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition*. J Nonlinear Sci 25, 1307–1346 (2015).

Korda, M., Mezić, *I. On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator.* J Nonlinear Sci 28, 687–710 (2018).

Klus S, Nüske F, Peitz S, Niemann J-H, Clementi C, Schütte C. *Data-driven approximation of the Koopman generator: Model reduction, system identification, and control.* Physica D: Nonlinear Phenomena. 2020;406.

Kernel methods: Klus, S.; Nüske, F.; Hamzi, B. *Kernel-Based Approximation of the Koopman Generator and Schrödinger Operator*. Entropy 2020, 22, 722.

J. Nathan Kutz, J. L. Proctor, S. L. Brunton, *Applied Koopman Theory for Partial Di erential Equations and Data-Driven Modeling of Spatio-Temporal Systems*, Complexity, vol. 2018, Article ID 6010634, 16 pages, 2018.

Control applications: Milan Korda, Igor Mezić, *Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control*, Automatica, Volume 93, 2018, Pages 149-160.

With deep learning: Yeung, E., Kundu, S., Hodas, N.O. (2019). *Learning Deep Neural Network Representations for Koopman Operators of Nonlinear Dynamical Systems.* 2019 American Control Conference (ACC), 4832-4839.

Ping, Z, Yin, Z, Li, X, Liu, Y, Yang, T. *Deep Koopman model predictive control for enhancing transient stability in power grids.* Int J Robust Nonlinear Control. 2020; 1– 15.