Koopman Kernel Regression

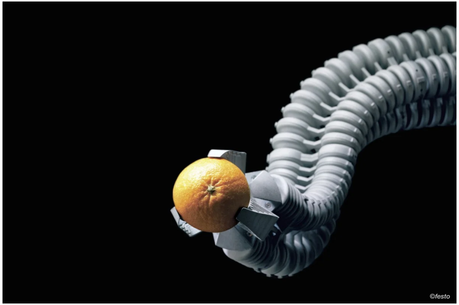
Petar Bevanda[1] Max Beier[1] Armin Lederer[1] Stefan Sosnowski[1]
Eyke Hüllermeier[2] Sandra Hirche[1]

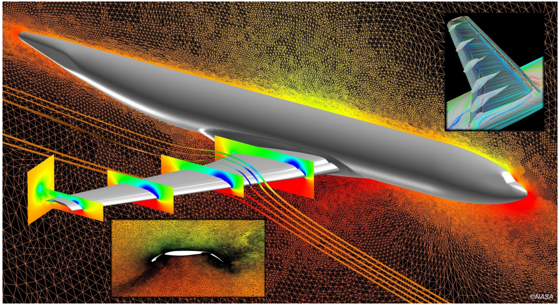
[1]Chair of Information-oriented Control
Technical University of Munich

[2]Chair of Artificial Intelligence and Machine Learning
Ludwig Maximilian University of Munich

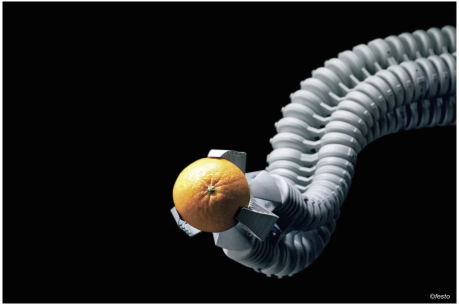
37th Conference on Neural Information Processing Systems (NeurIPS 2023)

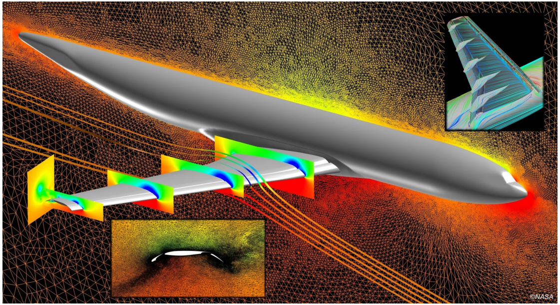
Simple yet Expressive Representations of Complex Dynamics





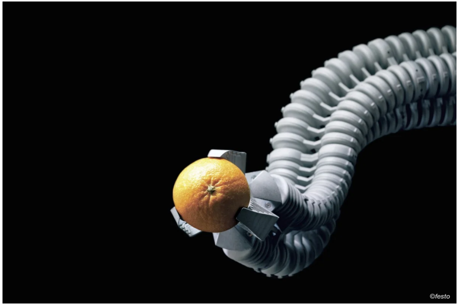
Simple yet Expressive Representations of Complex Dynamics

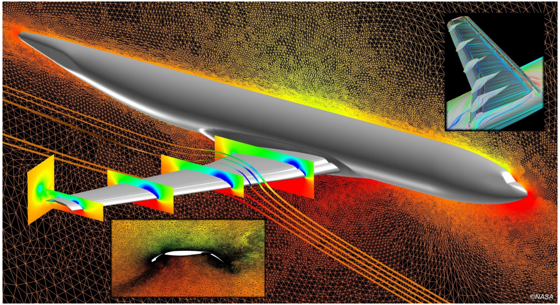




simple & universal **representations**

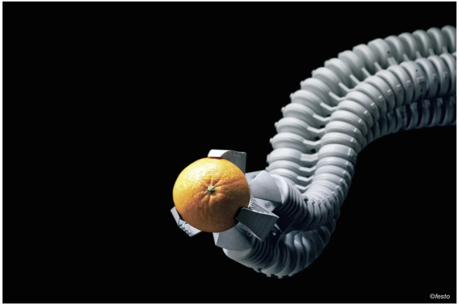
Simple yet Expressive Representations of Complex Dynamics

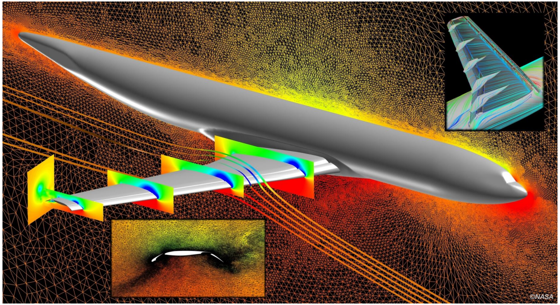




simple & universal **representations**
guaranteed **generalization** & **consistency**

Simple yet Expressive Representations of Complex Dynamics





simple & universal **representations**
guaranteed **generalization** & **consistency**

) *facilitating accelerated optimization-based decision making*

Take the best of time-series, ODEs and RKHS

Time-series decomposition

discriminative3
linear 3
time-variant 7

ODEs

generative 3
nonlinear7
time-invariant 3

Take the best of time-series, ODEs and RKHS

Time-series decomposition

discriminative3
linear 3
time-variant 7

ODEs

generative 3
nonlinear7
time-invariant 3

Learning in the best of both worlds?

responses2 span(time-invariant behaviors) 2 RKHS

Linear & Dynamics-invariant RKHS for Learning Dynamics

Koopmanism

Symbol $\mathcal{K}$: the time-shift or Koopman operator, so $\mathcal{K}y(k) = y(k+1)$.

Linear & Dynamics-invariant RKHS for Learning Dynamics

Koopmanism

Symbol K: the time-shift or Koopman operator, so $K y(k) = y(k+1)$.

Koopman eigenfunctions

$$K^k \;\; _j = \;\; ^k_j \;\; _j$$

! evolve in time with amplitude $| \;\; _j |$ & frequency $\arg(\;\; _j)$

Linear & Dynamics-invariant RKHS for Learning Dynamics

Koopmanism

Symbol K: the time-shift or Koopman operator, so $K y(k) = y(k+1)$.

Koopman eigenfunctions

$$K^k \;{}_j = \;{}^k_j \;{}_j$$

! evolve in time with amplitude $|\;{}_j|$ & frequency $\arg(\;{}_j)$

Ours

invariant RKHSs $\{H\;{}_j\}$

Existing

[Kostic+ 2022; Klus+ 2020]

Koopman Kernel Regression

Koopman kernel (ridge) regression

Given N i.i.d. state-output trajectories $\{x^{(i)}, y^{(i)}\}_{i=1}^{N} \in (X \times Y)^N$ of length H, compute

$$\hat{M} = \underset{M \in S(H_1 \to H_D)}{\arg\min} \left(\frac{1}{N} \sum_{i \in [N]} \| y^{(i)} - M(x^{(i)}) \|_Y^2 + \| M \|_H^2 \right) \in \mathrm{span}\{ K(\cdot\, ; x^{(i)}) \}$$

Koopman kernel

Koopman Kernel Regression

Koopman kernel (ridge) regression

Given N i.i.d. state-output trajectories $\{x^{(i)}, y^{(i)}\}_{i=1}^{N} \in (\mathbb{X} \times \mathbb{Y})^N$ of length H, compute

$$\hat{M} = \underset{M \in S(H_1 \otimes \cdots \otimes H_D)}{\arg\min} \left(\frac{1}{N} \sum_{i \in [N]} \| y^{(i)} - M(x^{(i)}) \|_{\mathbb{Y}}^2 + \lambda \| M \|_{H}^2 \right) \in \mathrm{span}\{ K(\cdot, x^{(i)}) \}$$

Koopman kernel

minimizes multi-step empirical risk

Koopman Kernel Regression

Koopman kernel (ridge) regression

Given N i.i.d. state-output trajectories $\{x^{(i)}, y^{(i)}\}_{i=1}^{N} \in (\mathcal{X} \times \mathcal{Y})^N$ of length H, compute

Koopman kernel

$$\hat{M} = \underset{M \in S(\mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_D)}{\arg\min} \left(\frac{1}{N} \sum_{i \in [N]} \| y^{(i)} - M(x^{(i)}) \|_{\mathcal{Y}}^2 + \lambda \| M \|_{\mathcal{H}}^2 \right) \in \operatorname{span}\{ K(\cdot, x^{(i)}) \}$$

minimizes multi-step empirical risk

approximation error vanishing with rank D

Koopman Kernel Regression

Koopman kernel (ridge) regression

Given N i.i.d. state-output trajectories $\{x^{(i)}, y^{(i)}\}_{i=1}^{N} \in (\mathbb{X} \times \mathbb{Y})^N$ of length H, compute

$$\hat{M} = \arg\min_{M \in S(\mathcal{H}_1 \cdots \mathcal{H}_D)} \left(\frac{1}{N} \sum_{i \in [N]} \| y^{(i)} - M(x^{(i)}) \|_{\mathcal{Y}}^2 + \lambda \| M \|_{\mathcal{H}}^2 \right) \in \operatorname{span}\{ K(\cdot, x^{(i)}) \}$$

Koopman kernel

minimizes multi-step empirical risk

approximation error vanishing with rank D

rank-independent generalization

Practical Implications

Flow past a cylinder

Forecasting velocity magnitude of sensor from 50 100-dimensional initial conditions.

Practical Implications

Flow past a cylinder

Forecasting velocity magnitude of sensor from 50 100-dimensional initial conditions.

Orders-of-magnitude greater accuracy (also for a wide range of hyperparameters)

Also in the Full Paper...

Including but not limited to

- full theoretical results
- numerical validation
- extensive comparisons
- complexity analysis
- forecasting and training times

Also in the Full Paper...

Including but not limited to

- full theoretical results
- numerical validation
- extensive comparisons
- complexity analysis
- forecasting and training times

Spoiler: superior to *Koopman operator regression* across the board

References

Bernard Haasdonk and Hans Burkhardt. **Invariant kernel functions for pattern analysis and machine learning**. In: *Machine Learning* 68.1 (July 2007), pp. 35–61. ISSN: 08856125. DOI: 10.1007/s10994-007-5009-7.

Stefan Klus, Ingmar Schuster and Krikamol Muandet. **Eigendecompositions of Transfer Operators in Reproducing Kernel Hilbert Spaces**. In: *Journal of Nonlinear Science* 30.1 (2020), pp. 283–315. ISSN: 1432-1467. DOI: 10.1007/s00332-019-09574-z.

Vladimir Kostic, Pietro Novelli, Andreas Maurer, Carlo Ciliberto, Lorenzo Rosasco and Massimiliano Pontil. **Learning Dynamical Systems via Koopman Operator Regression in Reproducing Kernel Hilbert Spaces**. In: *Advances in Neural Information Processing Systems*. 2022, pp. 4017–4031.

Cristopher Salvi, Thomas Cass, James Foster, Terry Lyons and Weixin Yang. **The Signature Kernel Is the Solution of a Goursat PDE**. In: *SIAM Journal on Mathematics of Data Science* 3.3 (Jan. 2021), pp. 873–899. ISSN: 2577-0187. DOI: 10.1137/20M1366794.