Representer Theorems for Learning Koopman Operators

Mohammad Khosravi

Delft Center for Systems and Control

Delft University of Technology

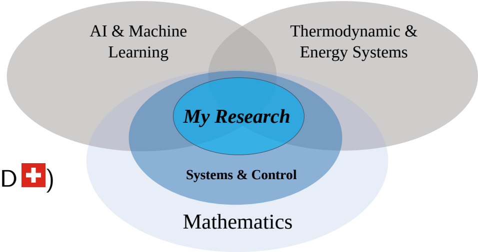
ERNSI 2023

KTH, Stockholm, Sweden

September 27th, 2023

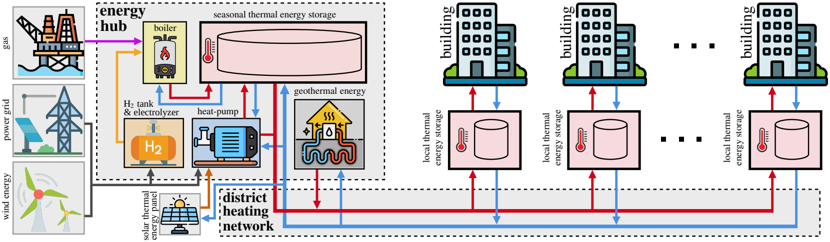
- Assistant Professor
  Delft Center for Systems & Control,

  TU Delft



- Systems & Control (BSc, MSc, PhD)
- Mathematics (BSc, MSc)

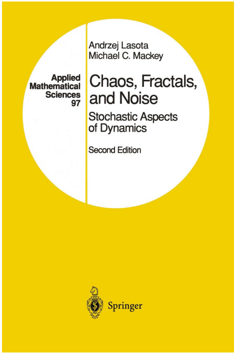
- **Theory**: {*Statistical Learning Theory*} $\cap$ {*Systems & Control*}
  $\rightsquigarrow$ reliable and efficient data-driven techniques
- **Application**: Smart & Safe Efficient Energy Systems $\rightsquigarrow$ net-zero CO_2 emission

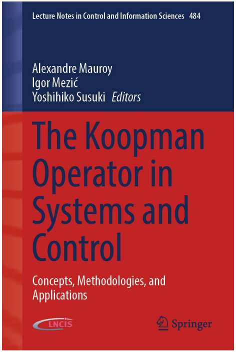


Koopman Operator

– Koopman operators appear in theory and practice of

- fractals and chaos
- fluid dynamics and PDEs
- nonlinear dynamics
- nonlinear control, e.g., in robotics
- neural networks and DL
- ...

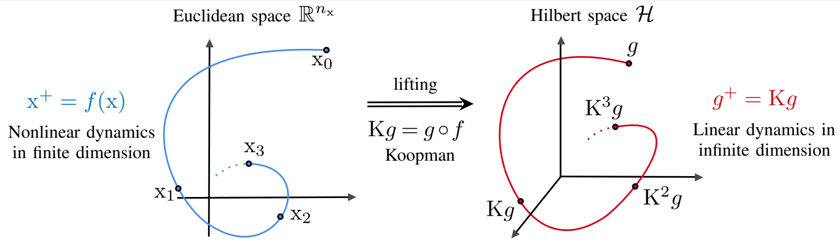




– **This talk**: Can we **learn** Koopman operators *efficiently*?

– Main reference:

M. Khosravi, "*Representer theorem for learning Koopman operators*", in IEEE Transactions on Automatic Control, vol. 68, no. 5, pp. 2995–3010, May 2023.



- Given space $\mathcal{X}$ and vector field: $f : \mathcal{X} \to \mathcal{X}$, we have a dynamical system as

$$x^+ = f(x)$$

- Hilbert space of observables: $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}} := \{g : \mathcal{X} \to \mathbb{R} \,|\, g \text{ measurable}\}$
- **Koopman operator**

$$\begin{aligned} K : \mathcal{H} &\longrightarrow \mathcal{H} \\ g(\cdot) &\longmapsto g(f(\cdot)) \qquad (\text{or } Kg = g \circ f) \end{aligned}$$

- Koopman operator $K \in \mathcal{L}(\mathcal{H})$ defines a **linear dynamical system** on $\mathcal{H}$

$$g^+ = K g$$

- **Main feature**: $(K^n g)(\cdot) = g(\underbrace{f(f(\cdots(f}_{n \text{ times}}(\cdot))\cdots)))$

- A trajectory of system: $\mathrm{x}_0, \mathrm{x}_1, \ldots, \mathrm{x}_{n_\mathrm{s}} \in \mathcal{X}$
- A set of observables: $g_1, g_2, \ldots, g_{n_\mathrm{g}} \in \mathcal{H}$
- We are given

$$\begin{array}{ccccc} g_1(\mathrm{x}_0), & g_1(\mathrm{x}_1), & g_1(\mathrm{x}_2), & \ldots & g_1(\mathrm{x}_{n_\mathrm{s}}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_l(\mathrm{x}_0), & g_l(\mathrm{x}_1), & g_l(\mathrm{x}_2), & \ldots & g_l(\mathrm{x}_{n_\mathrm{s}}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{n_\mathrm{g}}(\mathrm{x}_0), & g_{n_\mathrm{g}}(\mathrm{x}_1), & g_{n_\mathrm{g}}(\mathrm{x}_2), & \ldots & g_{n_\mathrm{g}}(\mathrm{x}_{n_\mathrm{s}}) \end{array}$$

$\rightsquigarrow$ trajectory $\mathrm{x}_0, \ldots, \mathrm{x}_{n_\mathrm{s}}$ through the lens of observable $g_l : \mathcal{X} \to \mathbb{R}$

- Data set: $\mathscr{D} := \left\{ y_{kl} := g_l(\mathrm{x}_k) \,\middle|\, k = 0, \ldots, n_\mathrm{s}, l = 1, \ldots, n_\mathrm{g} \right\}$,
  or $\mathscr{D} := \left\{ (\mathrm{x}_k, y_{kl} := g_l(\mathrm{x}_k)) \,\middle|\, k = 0, \ldots, n_\mathrm{s}, l = 1, \ldots, n_\mathrm{g} \right\}$

Problem (Learning Koopman Operator)

Learn Koopman operator $\mathrm{K} \in \mathcal{L}(\mathcal{H})$ *using data* $\mathcal{D}$*, i.e., find* $\hat{\mathrm{K}}$ *such that* $\hat{\mathrm{K}}g \approx g \circ f$.

- Drawbacks of existing methods and problem formulations:
  - *Not General*
  - *Not Rigorous* – ad hoc and imprecise formulation
  - *Indirect Approach* – structural approximation + parameter estimation
  - *Expert-type Knowledge* – eigenfunctions are given!

Problem (Learning Koopman Operator)

Learn Koopman operator $\mathrm{K} \in \mathcal{L}(\mathcal{H})$ *using data* $\mathcal{D}$*, i.e., find* $\hat{\mathrm{K}}$ *such that* $\hat{\mathrm{K}}g \approx g \circ f$.

- Generic form of Koopman operator learning problem:

$$\begin{aligned} \min_{\mathrm{K}\in\mathcal{L}(\mathcal{H})} \quad & \mathcal{E}(\mathrm{K}) + \lambda\mathcal{R}(\mathrm{K}) \\ \text{s.t.} \quad & \mathrm{K} \in \mathcal{C} \end{aligned} \qquad (\star)$$

- Main ingredients of the problem:
  - empirical loss $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ ⇝ to evaluate fitting on the data
  - regularization $\mathcal{R} : \mathcal{L}(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ ⇝ to avoid overfitting and/or for soft incorporation of **side-information**
  - constraints $\mathcal{C} \subset \mathcal{L}(\mathcal{H})$ ⇝ other constraints of interest or hard incorporation of **side-information**
- **Question:** Considering that $(\star)$ is an **infinite-dimensional optimization** problem, **when** and **how** can ve solve it?

- Learning Koopman operator with Tikhonov regularization

$$\min_{\mathrm{K}\in\mathcal{L}(\mathcal{H})} \sum_{k=1}^{n_{\mathrm{s}}}\sum_{l=1}^{n_{\mathrm{g}}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\big)^2 + \lambda\|\mathrm{K}\|^2 \qquad (\dagger)$$

- The **evaluation operator** at $\mathrm{x}\in\mathcal{X}$

$$\begin{aligned}\mathrm{e}_{\mathrm{x}}:\quad &\mathcal{H}\to\mathbb{R}\\ &g\mapsto g(\mathrm{x})\end{aligned}$$

(recall that $g:\mathcal{X}\to\mathbb{R}$)

- **Assumption I**: The evaluation operators $\mathrm{e}_{\mathrm{x}_0},\ldots,\mathrm{e}_{\mathrm{x}_{n_{\mathrm{s}}-1}}$ are **bounded**, i.e., there exists $v_1,\ldots,v_{n_{\mathrm{s}}}$ such that

$$\langle v_k, g\rangle = \mathrm{e}_{\mathrm{x}_{k-1}}(g) = g(\mathrm{x}_{k-1}),\qquad \text{for } k=1,\ldots,n_{\mathrm{s}}, \text{ and any } g\in\mathcal{H}$$

- $\mathrm{V} := \big[\langle v_{k_1}, v_{k_2}\rangle\big]_{k_1,k_2=1}^{n_{\mathrm{s}}}$ the **Gram matrix** of $v_1,\ldots,v_{n_{\mathrm{s}}}$
- $\mathrm{G} := \big[\langle g_{l_1}, g_{l_2}\rangle\big]_{l_1,l_2=1}^{n_{\mathrm{g}}}$ the **Gram matrix** of $g_1,\ldots,g_{n_{\mathrm{g}}}$
- $\mathrm{Y} := \big[y_{kl}\big]_{k=1,l=1}^{n_{\mathrm{s}},\,n_{\mathrm{g}}} = \big[g_l(\mathrm{x}_k)\big]_{k=1,l=1}^{n_{\mathrm{s}},\,n_{\mathrm{g}}}$

Theorem 1

– *Under Assumption I, the learning problem*

$$\min_{\mathrm{K}\in\mathcal{L}(\mathcal{H})} \sum_{k=1}^{n_{\mathrm{s}}}\sum_{l=1}^{n_{\mathrm{g}}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\big)^2 + \lambda\|\mathrm{K}\|^2 \qquad (\dagger)$$

has a ***unique*** *solution.*

– *The solution of* $(\dagger)$ admits a **parametric representation** as

$$\hat{\mathrm{K}} = \sum_{k=1}^{n_{\mathrm{s}}}\sum_{l=1}^{n_{\mathrm{g}}} a_{kl}\ v_k \otimes g_l$$

– The coefficient matrix $\mathrm{A} = [a_{kl}]_{k=1,l=1}^{n_{\mathrm{s}},\,n_{\mathrm{g}}} \in \mathbb{R}^{n_{\mathrm{s}}\times n_{\mathrm{g}}}$ is the solution of following **finite-dimensional convex optimization problem**

$$\min_{\mathrm{A}\in\mathbb{R}^{n_{\mathrm{s}}\times n_{\mathrm{g}}}} \|\mathrm{V}\mathrm{A}\mathrm{G} - \mathrm{Y}\|_{\mathrm{F}}^2 + \lambda\,\|\mathrm{V}^{\frac{1}{2}}\mathrm{A}\mathrm{G}^{\frac{1}{2}}\|^2$$

TU Delft

- Given $\mathcal{W}$, a closed linear subspace of $\mathcal{H}$, define Koopman learning problem as

$$\begin{aligned} \min_{\mathrm{K}\in\mathcal{L}(\mathcal{H})} \quad & \sum_{k=1}^{n_\mathrm{s}}\sum_{l=1}^{n_\mathrm{g}}\Big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\Big)^2 + \lambda\|\mathrm{K}\|^2 \\ \text{s.t.} \quad & \mathrm{K} \in \mathcal{L}_{\mathcal{W}} := \Big\{\mathrm{S}\in\mathcal{L}(\mathcal{H}) \;\Big|\; \mathrm{S}(g_l)\in\mathcal{W},\ \text{for } l=1,\ldots,n_\mathrm{g}\Big\} \end{aligned} \qquad (\ddagger)$$

Theorem 2

– *Under Assumption I, the learning problem* $(\ddagger)$ has a **unique** solution. The solution of $(\ddagger)$ admits a **parametric representation** as

$$\hat{\mathrm{K}} = \sum_{k=1}^{n_\mathrm{s}}\sum_{l=1}^{n_\mathrm{g}} a_{kl}\ (\Pi_{\mathcal{W}}v_k)\otimes g_l$$

– The coefficient matrix $\mathrm{A} = [a_{kl}]_{k=1,l=1}^{n_\mathrm{s},\,n_\mathrm{g}} \in \mathbb{R}^{n_\mathrm{s}\times n_\mathrm{g}}$ is the solution of following **finite-dimensional convex optimization problem**

$$\min_{\mathrm{A}\in\mathbb{R}^{n_\mathrm{s}\times n_\mathrm{g}}} \ \|\mathrm{W}_{\mathcal{V}}\mathrm{A}\mathrm{G}-\mathrm{Y}\|_\mathrm{F}^2 + \lambda\,\|\mathrm{W}_{\mathcal{V}}{}^{\frac{1}{2}}\mathrm{A}\mathrm{G}^{\frac{1}{2}}\|^2$$

where $\mathrm{W}_{\mathcal{V}}$ is the Gram matrix of $\Pi_{\mathcal{W}}v_1,\ldots,\Pi_{\mathcal{W}}v_{n_\mathrm{s}}$.

- EDMD approximate K by a finite-dimensional map $\mathrm{U} : \mathcal{G} \to \mathcal{G}$, where

$$\mathcal{G} := \mathrm{span}\{g_1, \ldots, g_{n_\mathrm{g}}\}.$$

  So, U has a **matrix** representation in basis $\{g_1, \ldots, g_{n_\mathrm{g}}\}$ as

$$\mathrm{U} g_l = \sum_{j=1}^{n_\mathrm{g}} [\mathrm{M}]_{(j,l)} g_j, \quad \text{for } l = 1, \ldots, n_\mathrm{g}$$

- EDMD employs data to estimate M as

$$\mathrm{M}^\star = \operatorname*{argmin}_{\mathrm{M} \in \mathbb{R}^{n_\mathrm{g} \times n_\mathrm{g}}} \sum_{k=1}^{n_\mathrm{s}} \sum_{j=1}^{n_\mathrm{g}} \Big((\mathrm{U} g_j)(\mathrm{x}_{k-1}) - g_j(\mathrm{x}_k) \Big)^2$$

- Replacing $\mathcal{W}$ with $\mathcal{G}$ in our approach, the Koopman learning problem becomes

$$\begin{array}{cl} \min\limits_{\mathrm{K} \in \mathcal{L}(\mathcal{H})} & \displaystyle\sum_{k=1}^{n_\mathrm{s}} \sum_{l=1}^{n_\mathrm{g}} \Big(y_{kl} - (\mathrm{K} g_l)(\mathrm{x}_{k-1}) \Big)^2 + \lambda \|\mathrm{K}\|^2 \\ \text{s.t.} & \mathrm{K} \in \mathcal{L}_\mathcal{G} \end{array}$$

  where $\mathrm{K} \in \mathcal{L}_\mathcal{G} := \Big\{ \mathrm{S} \in \mathcal{L}(\mathcal{H}) \;\Big|\; \mathrm{S}(g_l) \in \mathcal{G}, \text{ for } l = 1, \ldots, n_\mathrm{g} \Big\}.$

Theorem 3

Define matrix $\mathrm{C}_{\mathrm{M}^\star}$ *as* $\mathrm{M}^\star\mathrm{G}^{-1}$, *and, let operator* $\hat{\mathrm{K}}_{\mathrm{U}} \in \mathcal{L}(\mathcal{H})$ *be defined as*

$$\hat{\mathrm{K}}_{\mathrm{U}} = \sum_{j=1}^{n_{\mathrm{g}}} \sum_{l=1}^{n_{\mathrm{g}}} [\mathrm{C}_{\mathrm{M}^\star}]_{(j,l)}\, g_j \otimes g_l.$$

Then, we have

$$\hat{\mathrm{K}}_{\mathrm{U}} \;\in\; \operatorname*{argmin}_{\mathrm{K}\in\mathcal{L}_{\mathcal{G}}} \sum_{k=1}^{n_{\mathrm{s}}} \sum_{l=1}^{n_{\mathrm{g}}} \Big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\Big)^2.$$

Also, the EDMD map U *coincides with the restriction of* $\hat{\mathrm{K}}_{\mathrm{U}}$ *to* $\mathcal{G}$, *i.e.,* $\mathrm{U} = \hat{\mathrm{K}}_{\mathrm{U}}\big|_{\mathcal{G}}$.

Theorem 4

For $\lambda > 0$, *let* $\hat{\mathrm{K}}_\lambda$ *be the unique solution of*

$$\min_{\mathrm{K}\in\mathcal{L}_{\mathcal{G}}} \mathcal{J}_\lambda(\mathrm{K}) := \sum_{k=1}^{n_{\mathrm{s}}} \sum_{l=1}^{n_{\mathrm{g}}} \Big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\Big)^2 + \lambda\|\mathrm{K}\|^2.$$

Then, $\lim_{\lambda\downarrow 0} \hat{\mathrm{K}}_\lambda = \hat{\mathrm{K}}_{\mathrm{U}}$ *and* $\lim_{\lambda\to\infty} \hat{\mathrm{K}}_\lambda = 0$, *both in operator norm topology.*

- Generic form of Koopman operator learning problem:

$$\begin{aligned} \min_{\mathrm{K}\in\mathcal{F}} \quad & \mathcal{E}(\mathrm{K}) + \lambda\mathcal{R}(\mathrm{K}) \\ \text{s.t.} \quad & \mathrm{K}\in\mathcal{C} \end{aligned}$$

- Ingredients of the problem:
  - empirical loss $\mathcal{E} : \mathcal{L}(\mathcal{H}) \to \mathbb{R}\cup\{+\infty\}$
  - regularization term $\mathcal{R} : \mathcal{L}(\mathcal{H}) \to \mathbb{R}\cup\{+\infty\}$
  - constraint set $\mathcal{C} \subseteq \mathcal{L}(\mathcal{H})$
  - **regularization weight** $\lambda > 0$
  - **feasible set** $\mathcal{F} = \mathcal{L}(\mathcal{H})$ or $\mathcal{L}_{\mathcal{W}}$, where $\mathcal{W}$ is a closed linear subspace of $\mathcal{H}$.

  Recall that: $\mathcal{L}_{\mathcal{W}} = \left\{\mathrm{S}\in\mathcal{L}(\mathcal{H}) \;\middle|\; \mathrm{S}(g_l)\in\mathcal{W},\ \text{for } l=1,\ldots,n_{\mathrm{g}}\right\}$
  representer vectors $z_k := v_k$ or $\Pi_{\mathcal{W}}v_k$, for $k=1,\ldots,n_{\mathrm{s}}$
  define $\mathcal{Z} := \operatorname{span}\{z_1,\ldots,z_{n_{\mathrm{s}}}\}$
  define Z as Gram matrix of $z_1,\ldots,z_{n_{\mathrm{s}}}$

- Consider index set $\mathcal{I} \subseteq \{1, \ldots, n_{\mathrm{s}}\} \times \{1, \ldots, n_{\mathrm{g}}\}$.
- Define vector $\mathrm{y}_{\mathcal{I}} := \big[y_{kl}\big]_{(k,l)\in\mathcal{I}} = \big[g_l(\mathrm{x}_k)\big]_{(k,l)\in\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|}$.
- Let $\ell : \mathbb{R}^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{I}|} \to \mathbb{R} \cup \{+\infty\}$ be a function such that

$$\ell(\cdot, \mathrm{y}_{\mathcal{I}}) : \mathbb{R}^{|\mathcal{I}|} \to \mathbb{R} \cup \{+\infty\}$$

  is a proper **convex** function.
- The **generalized empirical loss** $\mathcal{E}_\ell : \mathcal{L}(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ can be defined in the following general form

$$\begin{aligned}\mathcal{E}_\ell(\mathrm{K}) :=& \ell\Big(\big[(\mathrm{K}g_l)(\mathrm{x}_{k-1})\big]_{(k,l)\in\mathcal{I}}, \mathrm{y}_{\mathcal{I}}\Big)\\ =& \ell\Big(\big[(\mathrm{K}g_l)(\mathrm{x}_{k-1})\big]_{(k,l)\in\mathcal{I}}, \big[y_{kl}\big]_{(k,l)\in\mathcal{I}}\Big)\end{aligned}$$

- Various forms of empirical loss are in this form, including

$$\mathcal{E}_\ell(\mathrm{K}) = \sum_{k=1}^{n_{\mathrm{s}}} \sum_{l=1}^{n_{\mathrm{g}}} \Big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\Big)^2$$

- **Example I**: empirical loss for **outlier robust regression**

$$\mathcal{E}_{\ell}(\mathrm{K}) \text{ or } \mathcal{E}_{L}(\mathrm{K}) := \sum_{k=1}^{n_{\mathrm{s}}}\sum_{l=1}^{n_{\mathrm{g}}} L\big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\big)^2,$$

where, given $\rho >$, loss function $L : \mathbb{R}_+ \to \mathbb{R}_+$ is defined as

  - Huber loss: $L(e) = \begin{cases} e^2, & \text{if } |e| \leq \rho \\ 2|e|\rho - \rho^2, & \text{otherwise} \end{cases}$
  - pseudo-Huber loss: $L(e) = (e^2 + \rho^2)^{\frac{1}{2}} - \rho$

- **Example II**: empirical loss based on **robust optimization**

$$\mathcal{E}_{\ell}(\mathrm{K}) \text{ or } \mathcal{E}_{\mathcal{U}}(\mathrm{K}) := \max_{\Delta\in\mathcal{U}} \sum_{k=1}^{n_{\mathrm{s}}}\sum_{l=1}^{n_{\mathrm{g}}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1}) - \Delta_{(k,l)}\big)^2$$

- **Example III**: empirical loss based on **distributionally robust optimization**

$$\mathcal{E}_{\ell}(\mathrm{K}) \text{ or } \mathcal{E}_{\mathcal{P}}(\mathrm{K}) := \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\Delta\sim\mathbb{P}} \left[\sum_{k=1}^{n_{\mathrm{s}}}\sum_{l=1}^{n_{\mathrm{g}}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1}) - \Delta_{(k,l)}\big)^2\right]$$

- Define **augmented regularization** $\overline{\mathcal{R}} : \mathcal{L}(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ as $\overline{\mathcal{R}} := \mathcal{R} + \delta_{\mathcal{C}}$
- **Assumption II**: For any $\mathrm{S} \in \mathcal{L}(\mathcal{H})$, we have $\overline{\mathcal{R}}(\Pi_{\mathcal{Z}}\,\mathrm{S}\,\Pi_{\mathcal{G}}) \leq \overline{\mathcal{R}}(\mathrm{S})$.

Theorem 5

– *Under the mentioned assumptions, if the learning problem*

$$\begin{array}{ll} \min\limits_{\mathrm{K}\in\mathcal{F}} & \mathcal{E}_\ell(\mathrm{K}) + \lambda\mathcal{R}(\mathrm{K}) \\ \text{s.t.} & \mathrm{K} \in \mathcal{C}, \end{array} \qquad (\star)$$

admits a solution, then it has a solution with following ***parametric representation***

$$\hat{\mathrm{K}} = \sum_{k=1}^{n_{\mathrm{z}}} \sum_{l=1}^{n_{\mathrm{g}}} a_{kl}\; z_k \otimes g_l$$

– *When $\mathcal{D} := \mathcal{F} \cap \mathrm{dom}(\mathcal{R}) \cap \mathcal{C}$ is a non-empty, closed and convex set, $\mathcal{R}$ is convex and lower semi-continuous, and $\overline{\mathcal{R}}$ is coercive, then $(\star)$* admits **at least one solution** with the given parametric representation.

– Additionally, if $\mathcal{R}$ is strictly convex on $\mathcal{D}$, then the solution of $(\star)$ is **unique**.

- Define **augmented regularization** $\overline{\mathcal{R}} : \mathcal{L}(\mathcal{H}) \to \mathbb{R} \cup \{+\infty\}$ as $\overline{\mathcal{R}} := \mathcal{R} + \delta_{\mathcal{C}}$
- **Assumption II**: For any $\mathrm{S} \in \mathcal{L}(\mathcal{H})$, we have $\overline{\mathcal{R}}(\Pi_{\mathcal{Z}}\, \mathrm{S}\, \Pi_{\mathcal{G}}) \leq \overline{\mathcal{R}}(\mathrm{S})$.

Theorem 6

Let $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_+$,

- $\mathcal{R}_i : \mathcal{L}(\mathcal{H}) \to \mathbb{R}_+ \cup \{+\infty\}$ *be a regularization function, for* $i = 1, \ldots, m$,
- $\mathcal{C}_\alpha \subseteq \mathcal{H}$, *for each* $\alpha \in \mathcal{A}$. *($\mathcal{A}$: arbitrary index set)*

Define

- $\mathcal{C} := \cap_{\alpha \in \mathcal{A}} \mathcal{C}_\alpha$ *and*
- $\overline{\mathcal{R}}_{i,\alpha} := \lambda_i \mathcal{R}_i + \delta_{\mathcal{C}_\alpha}$, *for* $\alpha \in \mathcal{A}$.

If Assumption II is satisfied by $\overline{\mathcal{R}}_{i,\alpha}$, *for each* $\alpha \in \mathcal{A}$ *and* $i = 1, \ldots, m$, *then* $\overline{\mathcal{R}} : \mathcal{L}(\mathcal{H}) \to \mathbb{R}_+ \cup \{+\infty\}$ *defined as*

$$\overline{\mathcal{R}} := \sum_{i=1}^{m} \lambda_i \mathcal{R}_i + \delta_{\mathcal{C}}$$

also satisfies Assumption II.

- Given orthonormal basis $\{b_k\}_{k=1}^{\infty}$, Frobenius norm of operator K

$$\|\mathrm{K}\|_{\mathrm{F}}^2 = \sum_{k=1}^{\infty} \langle b_k, \mathrm{K}^*\mathrm{K} b_k \rangle = \sum_{k=1}^{\infty} \|\mathrm{K} b_k\|^2$$

- The problem of learning Koopman operator with **Frobenius norm** regularization

$$\min_{\mathrm{K}\in\mathcal{F}} \ \mathcal{E}_\ell(\mathrm{K}) + \lambda\|\mathrm{K}\|_{\mathrm{F}}^2 \qquad (\dagger)$$

Theorem 7

– *Under previous assumptions and if $\lambda > 0$, the optimization problem* $(\dagger)$ admits a **unique** solution $\hat{\mathrm{K}}$ with parametric form

$$\hat{\mathrm{K}} = \sum_{k=1}^{n_{\mathrm{z}}} \sum_{l=1}^{n_{\mathrm{g}}} a_{kl} \ z_k \otimes g_l$$

– If $\mathcal{E}_\ell(\mathrm{K}) = \sum_{k=1}^{n_{\mathrm{s}}} \sum_{l=1}^{n_{\mathrm{g}}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\big)^2$, then $\mathrm{A} = [a_{kl}]_{k=1,l=1}^{n_{\mathrm{z}},\,n_{\mathrm{g}}}$ is the solution of following **finite-dimensional convex quadratic program**

$$\min_{\mathrm{A}\in\mathbb{R}^{n_{\mathrm{z}}\times n_{\mathrm{g}}}} \ \|\mathrm{ZAG} - \mathrm{Y}\|_{\mathrm{F}}^2 \ + \ \lambda\|\mathrm{Z}^{\frac{1}{2}}\mathrm{A}\mathrm{G}^{\frac{1}{2}}\|_{\mathrm{F}}^2 \qquad (\dagger')$$

- There is a **closed-form solution** for $(\dagger')$, and thus, for $(\dagger)$.

- The problem of learning Koopman operator with **rank constraint**

$$\begin{aligned} \min_{\mathrm{K}\in\mathcal{F}} \quad & \mathcal{E}_\ell(\mathrm{K}) \\ \text{s.t.} \quad & \mathrm{K} \in \mathcal{C} := \left\{ \mathrm{S}\in\mathcal{F} \,\middle|\, \mathrm{rank}(\mathrm{S}) := \dim(\mathrm{S}(\mathcal{H})) \le r \right\} \end{aligned} \qquad (\ddagger)$$

Theorem 8

Under previous assumptions, if the optimization problem $(\ddagger)$ admits a solution, then it has a solution $\hat{\mathrm{K}}$ with parametric form

$$\hat{\mathrm{K}} = \sum_{k=1}^{n_\mathrm{z}} \sum_{l=1}^{n_\mathrm{g}} a_{kl}\; z_k \otimes g_l$$

If $\mathcal{E}_\ell(\mathrm{K}) = \sum_{k=1}^{n_\mathrm{s}} \sum_{l=1}^{n_\mathrm{g}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\big)^2$, then $\mathrm{A} = [a_{kl}]_{k=1,l=1}^{n_\mathrm{z},\,n_\mathrm{g}}$ is the solution of following **finite-dimensional optimization** problem

$$\begin{aligned} \min_{\mathrm{A}\in\mathbb{R}^{n_\mathrm{z}\times n_\mathrm{g}}} \quad & \|\mathrm{ZAG}-\mathrm{Y}\|_\mathrm{F}^2, \\ \text{s.t.} \quad & \mathrm{rank}(\mathrm{ZAG}) \le r \end{aligned} \qquad (\ddagger')$$

- Using *Eckart-Young-Mirsky Theorem*, we can solve $(\ddagger')$, and subsequently, $(\ddagger)$.

- Nuclear norm of operator K

$$\|\mathrm{K}\|_* = \sup \left\{ |\mathrm{tr}(\mathrm{CK})| \,\middle|\, \mathrm{C} \text{ compact}, \ \|\mathrm{C}\| \leq 1 \right\}$$

- The problem of learning Koopman operator with **nuclear norm** regularization

$$\min_{\mathrm{K}\in\mathcal{F}} \ \mathcal{E}_\ell(\mathrm{K}) + \lambda \|\mathrm{K}\|_* \qquad (\star)$$

Theorem 9

– *Under previous assumptions and if $\lambda > 0$, the optimization problem* $(\star)$ admits a solution $\hat{\mathrm{K}}$ with parametric form

$$\hat{\mathrm{K}} = \sum_{k=1}^{n_\mathrm{z}} \sum_{l=1}^{n_\mathrm{g}} a_{kl} \ z_k \otimes g_l$$

– If $\mathcal{E}_\ell(\mathrm{K}) = \sum_{k=1}^{n_\mathrm{s}} \sum_{l=1}^{n_\mathrm{g}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\big)^2$, then $\mathrm{A} = [a_{kl}]_{k=1,l=1}^{n_\mathrm{z},\, n_\mathrm{g}}$ is the solution of following **finite-dimensional convex program**

$$\min_{\mathrm{A}\in\mathbb{R}^{n_\mathrm{z}\times n_\mathrm{g}}} \ \|\mathrm{ZAG} - \mathrm{Y}\|_\mathrm{F}^2 + \lambda \|\mathrm{Z}^{\frac{1}{2}}\mathrm{A}\mathrm{G}^{\frac{1}{2}}\|_*$$

Lemma 10

*Let $\mathrm{x}_{eq} = 0$ be an equilibrium point, i.e., $\mathrm{x}_{eq} = f(\mathrm{x}_{eq})$. Then, under certain conditions, x_{eq} is a **globally stable equilibrium point** if, for some $\varepsilon > 0$, we have*

$$\|\mathrm{K}\| \leq 1 - \varepsilon.$$

- The problem of learning **stable** Koopman operator

$$\begin{aligned} \min_{\mathrm{K} \in \mathcal{F}} \quad & \mathcal{E}_\ell(\mathrm{K}) + \lambda \mathcal{R}(\mathrm{K}) \\ \text{s.t.} \quad & \mathrm{K} \in \mathcal{C} \\ & \|\mathrm{K}\| \leq 1 - \varepsilon \end{aligned} \qquad (\star\star)$$

Theorem 11

Under previous assumptions, the existence, uniqueness and parametric representation

$$\hat{\mathrm{K}} = \sum_{k=1}^{n_\mathrm{z}} \sum_{l=1}^{n_\mathrm{g}} a_{kl} \; z_k \otimes g_l$$

are guaranteed for the solution of learning problem $(\star\star)$.

- Solving a **finite-dimensional convex program** with **quadratic objective** function and **LMI constraints** when
  i. $\mathcal{R}(\cdot) = \|\cdot\|^2$ or $\|\cdot\|_{\mathrm{F}}^2$
  ii. $\mathcal{E}_\ell(\mathrm{K}) = \sum_{k=1}^{n_\mathrm{s}} \sum_{l=1}^{n_\mathrm{g}} \big(y_{kl} - (\mathrm{K}g_l)(\mathrm{x}_{k-1})\big)^2$
  iii. no additional constrain

- The optimization problem
  - for the case of $\mathcal{R}(\cdot) = \|\cdot\|^2$

$$
\begin{array}{cl}
\min\limits_{\mathrm{B}\in\mathbb{R}^{n_\mathrm{z}\times n_\mathrm{g}},\,\beta\in\mathbb{R}} & \|\mathrm{Z}^{\frac{1}{2}}\mathrm{B}\mathrm{G}^{\frac{1}{2}} - \mathrm{Y}\|_{\mathrm{F}}^2 + \lambda\beta^2 \\
\text{s.t.} & \begin{bmatrix} \beta\mathbb{I}_{n_\mathrm{g}} & \mathrm{B} \\ \mathrm{B}^\mathsf{T} & \beta\mathbb{I}_{n_\mathrm{z}} \end{bmatrix} \succeq 0, \\
 & \beta \leq 1-\varepsilon,
\end{array}
$$

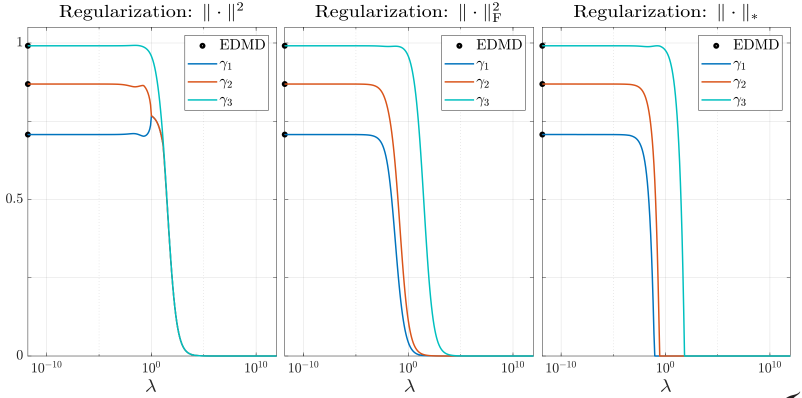
  - for the case of $\mathcal{R}(\cdot) = \|\cdot\|_{\mathrm{F}}^2$

$$
\begin{array}{cl}
\min\limits_{\mathrm{B}\in\mathbb{R}^{n_\mathrm{z}\times n_\mathrm{g}},\,\beta\in\mathbb{R}} & \|\mathrm{Z}^{\frac{1}{2}}\mathrm{B}\mathrm{G}^{\frac{1}{2}} - \mathrm{Y}\|_{\mathrm{F}}^2 + \lambda\|\mathrm{B}\|_{\mathrm{F}}^2 \\
\text{s.t.} & \begin{bmatrix} \beta\mathbb{I}_{n_\mathrm{g}} & \mathrm{B} \\ \mathrm{B}^\mathsf{T} & \beta\mathbb{I}_{n_\mathrm{z}} \end{bmatrix} \succeq 0, \\
 & \beta \leq 1-\varepsilon.
\end{array}
$$

- Nonlinear dynamical system

$$x_{1,k+1} = \mu_1 x_{1,k}$$
$$x_{2,k+1} = \mu_2 x_{2,k} + (\mu_1^2 - \mu_2) x_{1,k}^2$$

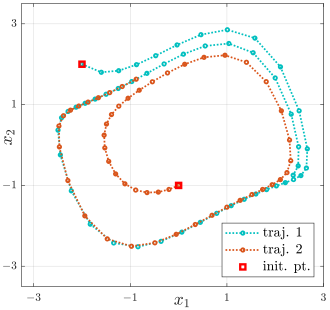
- Single trajectory, 60 samples, and **three** observables $\rightsquigarrow$ **three** eigenvalues
- Regularization term: $\|\cdot\|^2$, $\|\cdot\|_{\mathrm{F}}^2$, $\|\cdot\|_*$
- Regularization weight λ sweep: $0 \longleftrightarrow \infty$

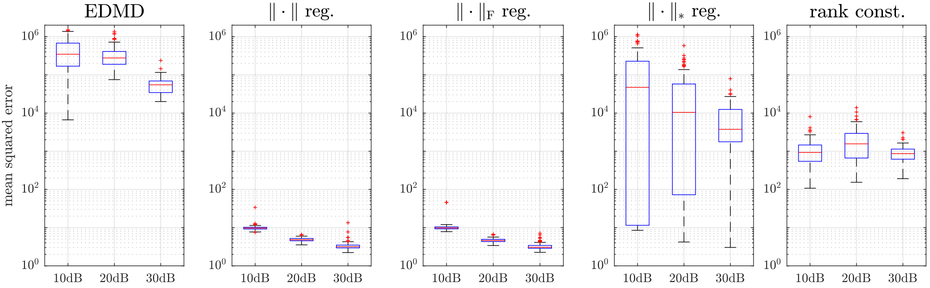


- Van der Pol oscillator

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = \mu(1 - x_1^2)x_2 - x_1$$

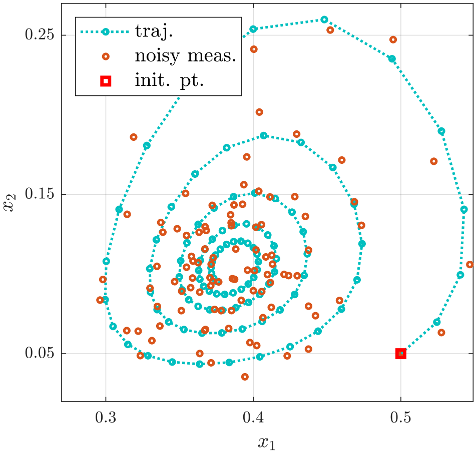
- Two trajectories
- Monte Carlo experiment: 120 noisy realizations
- SNR levels: 10 dB, 20 dB, 30 dB
- MSE on a set of test functions

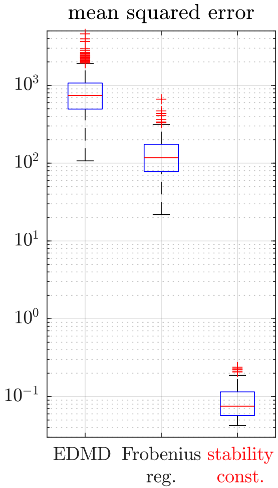




Dynamical system: Nicholson-Bailey model for host-parasitoid dynamics

Side-information: stability

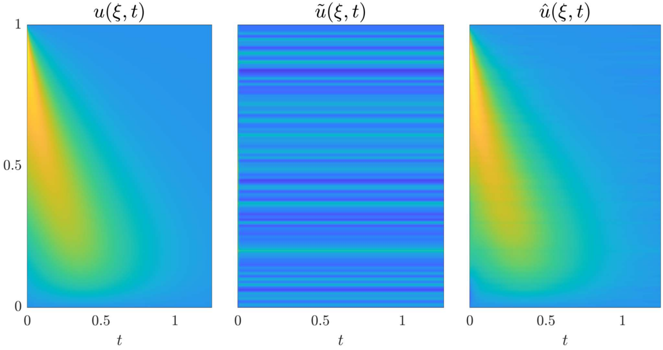




Dynamical system: advection-diffusion PDE

$$\begin{cases} \frac{\partial u}{\partial t}(\xi,t) = a\frac{\partial u}{\partial \xi}(\xi,t) + b\frac{\partial^2 u}{\partial \xi^2}(\xi,t), & (\xi,t) \in [0,1]\times[0,\infty) \\ u(\xi,0) = \sin(\pi\xi) & \xi \in [0,1] \end{cases}$$

New initial condition: $u(\xi,0) = 1 - \mathrm{e}^{-\xi}$, for $\xi \in [0,1]$



True EDMD Proposed

Conclusion:

- Though Koopman operators are **infinite-dimensional** objects, they can be learned efficiently, i.e.,
  - **directly from data**,
  - by solving **finite-dimensional convex programs**, and,
  - **without** *expert-type* knowledge.
- A **generalized representer theorem** holds for a wide range of empirical loss functions, regularizations, and constraints.
- **Side-information** incorporation can improve learning Koopman operators.

Outlook:

- Other candidates for regularization and constraints?
- Other forms of side-information?
- Generalizing the main results, e.g., to the case of **Banach** spaces or dynamics with **control** input?
- Applying the introduced learning approach to real data?
- Comparing with other Koopman operator learning methods?

Thanks for your attention!

Mohammad Khosravi

Mohammad.Khosravi@tudelft.nl
https://www.tudelft.nl/staff/mohammad.khosravi/
https://www.dcsc.tudelft.nl/~mohammadkhosravi/