2025 the 25th International Conference on Control, Automation, and Systems (ICCAS) @ Incheon, Korea | FrBT7

Necessary and Sufficient Conditions for Data-Driven Identification of Nonlinear Systems with Linear Embedding

Youngkyoung Kong, Inkyu Jang, and H. Jin Kim

Department of Aerospace Engineering,
Automation and Systems Research Institute (ASRI),
Seoul National University, Seoul, Korea

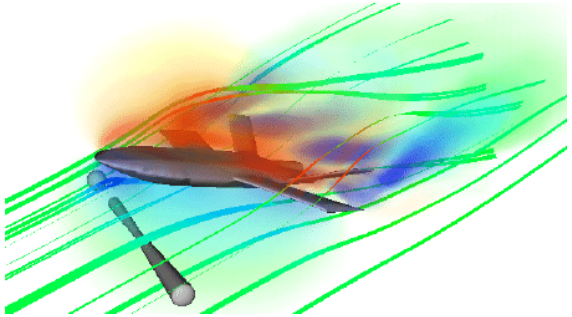
ICCAS 2025
The 25th International Conference on Control, Automation, and Systems

서울대학교
SEOUL NATIONAL UNIVERSITY

Motivation

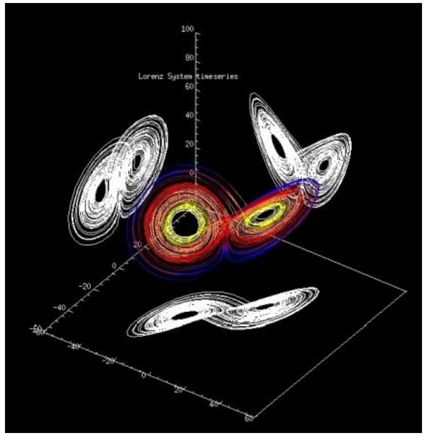
Modeling is Hard

- Accurately modeling dynamical systems is one of the fundamental problems in the field of control engineering.
- Several systems have known mathematical models. However, many real world systems…



too complex to model

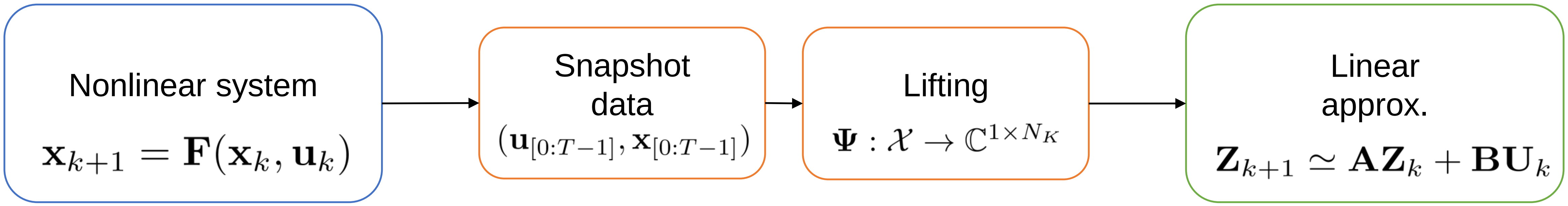
no accurate model



change unpredictably

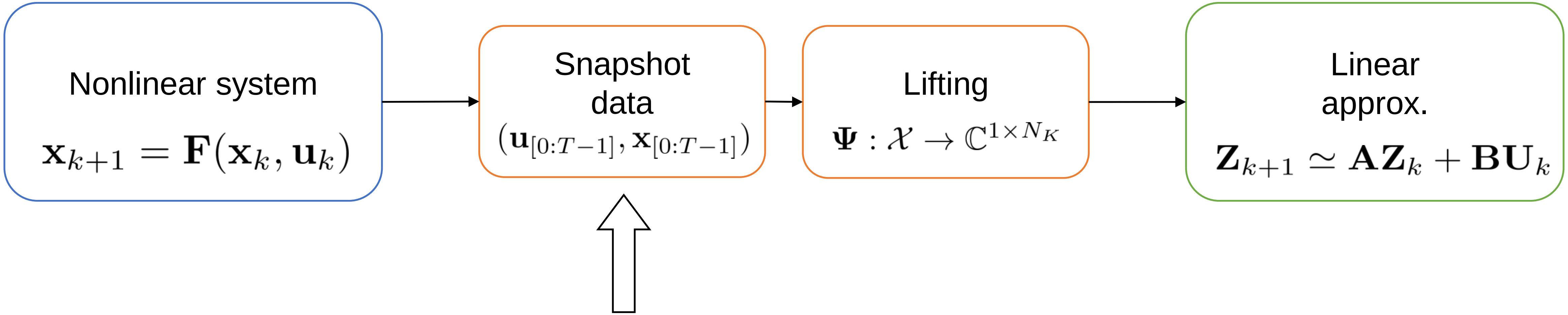
Data-Driven Modeling

- Koopman operator theory + Dynamic mode decomposition



Data-Driven Modeling

- Koopman operator theory + dynamic mode decomposition
  $\rightarrow$ do not provide conditions on
    1. how data should be collected,
    2. how much data is required to guarantee identifiability.



How to collect data efficiently?

Contributions

We adopt an identifiability condition from the behavioral systems theory.

- We establish a data collection condition by extending the rank condition on the Hankel matrix constructed from data.

- We theoretically derive the minimum number of data points required for system identification.

Koopman Operator Theory

Koopman Operator Theory

Definition 1 (Koopman operator). *Consider a discrete-time dynamical system $\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k)$ with $\mathbf{u}_k \equiv 0$, i.e., the autonomous system. For the function of state space $g \in \mathcal{F}$ with $g : \mathcal{X} \to \mathbb{C}$, which is called an observable, the Koopman operator $\mathcal{K} : \mathcal{F} \to \mathcal{F}$ is defined by*

$$\mathcal{K}g(\mathbf{x}) = g \circ \mathbf{F}(\mathbf{x}),$$

where $\mathcal{F}$ denotes a infinite-dimensional function space.

- There always exists a linear operator for every dynamical system that propagates its state on the infinite dimensional function space.

Extended Dynamic Mode Decomposition (EDMD)

- Infinite dimensional operator → Finite dimensional approx.

- In our paper, we only consider the nonlinear system that admits Koopman linear embedding.

Definition 2 (Koopman linear embedding). *The nonlinear system admits a Koopman linear embedding if there exists a basis vector of lifting functions*

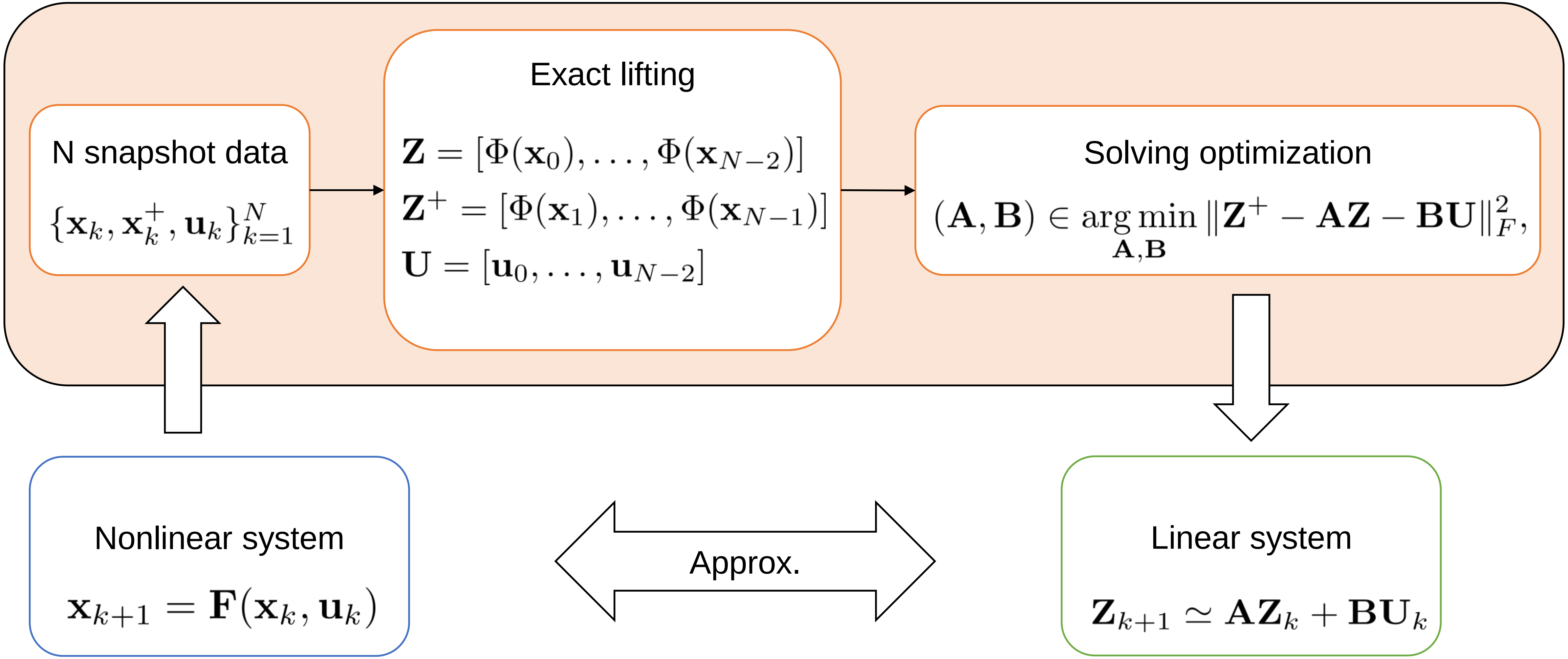
$$\mathbf{z}_k := \Phi(\mathbf{x}_k) = [\phi_1(\mathbf{x}_k), \phi_2(\mathbf{x}_k), \cdots, \phi_{n_z}(\mathbf{x}_k)]^\top \in \mathbb{R}^{n_z}$$

such that it evolves linearly along all trajectories. [1]

- There exists a set of the lifting functions that yield an exact linear representation.

[1] Shang, Xu, Jorge Cortés, and Yang Zheng. "Willems' fundamental lemma for nonlinear systems with Koopman linear embedding." *IEEE Control Systems Letters* (2024).

Extended Dynamic Mode Decomposition (EDMD)

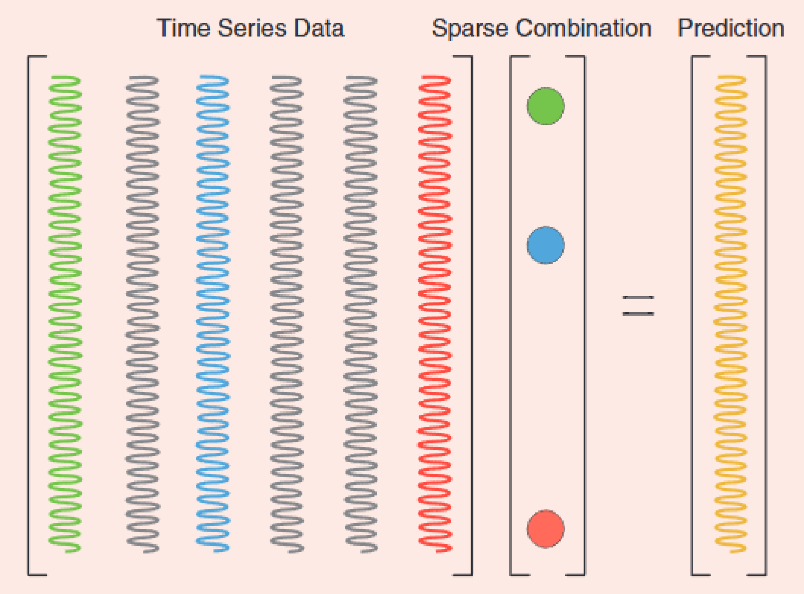


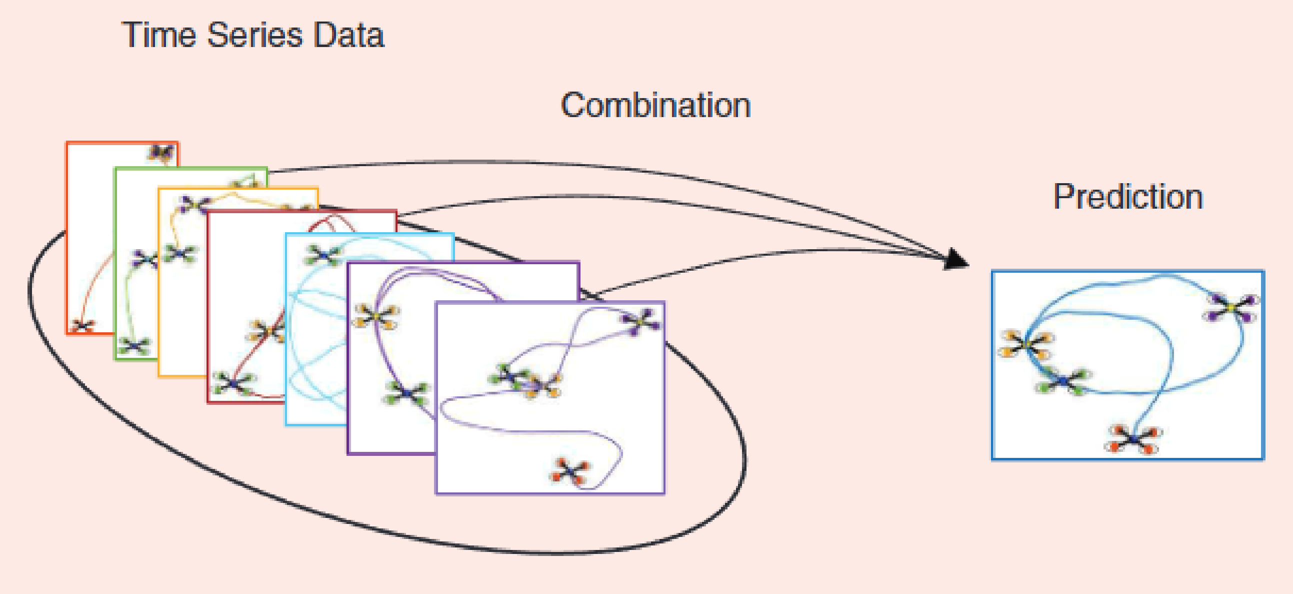
Behavioral Systems Theory

Behavioral Systems Theory

- Entire behavior space of a LTI system can be expressed as a linear combination of trajectories that sufficiently excite the system's modes.

  → If we collect data efficiently, we can always reconstruct the original LTI systems' trajectories from data.





[2] J. Coulson, "Jeremy Coulson [PhDs in Control]," in *IEEE Control Systems Magazine*, vol. 43, no. 3, pp. 132-134, June 2023, doi: 10.1109/MCS.2023.3253440. keywords: {Interviews},

Persistence of Excitation (PE)

Definition 3 (Persistently exciting)**.** *Consider a sequence of vector signals*

$$\mathbf{f}_{[1:T]} \triangleq \{\mathbf{f}_n\}_{n=1}^{T} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_T],$$

where $\mathbf{f} : \mathbb{Z} \to \mathbb{R}^{\bullet}$. *The sequence* $\mathbf{f}_{[1:T]}$ *is said to be persistently exciting of order* L *if its Hankel matrix of depth* L

$$\mathcal{H}_L(\mathbf{f}_{[1:T]}) = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 & \cdots & \mathbf{f}_{T-L+1} \\ \mathbf{f}_2 & \mathbf{f}_3 & \cdots & \mathbf{f}_{T-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{f}_L & \mathbf{f}_{L+1} & \cdots & \mathbf{f}_T \end{bmatrix}$$

has full row rank.

Willems' Fundamental Lemma

Lemma 1 (Willems' fundamental lemma). *Consider a LTI system, whose input-state-output trajectory tuple is $(\mathbf{u}_{[0:T-1]}, \mathbf{x}_{[0:T-1]}, \mathbf{y}_{[0:T-1]})$. Assume that the system is controllable and the input sequence $\mathbf{u}_{[0:T-1]}$ is persistently exciting of order $n_x + L$.*
Then, any input-output sequence $[\tilde{\mathbf{u}}^\top, \tilde{\mathbf{y}}^\top]^\top \in \mathbb{R}^{(n_u+n_y)L}$ is a valid trajectory of the system if and only if there exists $\mathbf{v} \in \mathbb{R}^{T-L+1}$ such that

$$\begin{bmatrix} \mathcal{H}_L(\mathbf{u}_{[0:T-1]}) \\ \mathcal{H}_L(\mathbf{y}_{[0:T-1]}) \end{bmatrix} \mathbf{v} = \begin{bmatrix} \tilde{\mathbf{u}} \\ \tilde{\mathbf{y}} \end{bmatrix}.$$

- If the input sequence is persistently exciting of order (state-dimension + L), then any valid output corresponding to a given input can be generated as a linear combination of past inputs and past outputs.

[3] Willems, Jan C., et al. "A note on persistency of excitation." *Systems & Control Letters* 54.4 (2005): 325-329.

Data-Driven Identifiability

Identifiability of Controllable LTI Systems

- Willems' fundamental lemma : only for controllable LTI systems!
- In general, lifted linear system is not controllable.
- We leverage the identifiability condition in [4].

[4] Markovsky, Ivan, and Florian Dörfler. "Identifiability in the behavioral setting." *IEEE Transactions on Automatic Control* 68.3 (2022): 1667-1677.

Identifiability of LTI systems

Theorem 2 (Necessary and sufficient conditions for identifiability of an LTI system). *Let the data $\mathscr{W}_d$ be generated by a system whose input dimension, state dimension and observability index are n_u, n_x, l, respectively. Then the system is identifiable from $\mathscr{W}_d$ if and only if*

$$\operatorname{rank} \mathcal{H}_{l+1}(\mathscr{W}_d) = n_u(l+1) + n_x. \quad [4]$$

[4] Markovsky, Ivan, and Florian Dörfler. "Identifiability in the behavioral setting." *IEEE Transactions on Automatic Control* 68.3 (2022): 1667-1677.

Partial Extension to Nonlinear Systems

Corollary 1 (Necessary and sufficient conditions for identifiability of a nonlinear system with linear embedding). *Assume the nonlinear system admits a Koopman linear embedding. Then the system is identifiable if and only if* $\operatorname{rank} \mathcal{H}_{l+1}(\mathscr{W}_d) = n_u(l+1) + n_z$ *holds for* $T \geq (n_u+1)(l+1) + n_z - 1$.

T : number of snapshots

n_z : lifted dimension

Partial Extension to Nonlinear Systems

$$\mathcal{H}_L = \underbrace{\left[\begin{array}{|c|c|c|c|} \hline \quad & \quad & \quad & \quad \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array}\right]}_{T - L + 1} \left.\vphantom{\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array}}\right\} L$$

$$\boxed{\begin{array}{l} T - L + 1 \overset{L = l + 1}{=} T - l \geq n_u(l + 1) + n_z \\[2em] \rightarrow \quad T \geq (n_u + 1)(l + 1) + n_z - 1 \end{array}}$$

Numerical Validations

System

$$\begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 0.99\, x_{1,k} \\ 0.9\, x_{2,k} + 1.2\, x_{1,k}^2 + 1.1\, x_{1,k}^3 + x_{1,k}^4 + 0.5\, u_k \end{bmatrix}$$

$$\Downarrow \quad \text{lifting} \quad \mathbf{z} := [x_1, x_2, x_1^2, x_1^3, x_1^4]^\top$$

$$\mathbf{z}_k = \begin{bmatrix} 0.99 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 1.2 & 1.1 & 1 \\ 0 & 0 & 0.99^2 & 0 & 0 \\ 0 & 0 & 0 & 0.99^3 & 0 \\ 0 & 0 & 0 & 0 & 0.99^4 \end{bmatrix} \mathbf{z}_k + \begin{bmatrix} 0 \\ 0.5 \\ 0 \\ 0 \\ 0 \end{bmatrix} u_k. \qquad \text{observability index}(l) : 4$$

Identification Condition

- rank condition : $\operatorname{rank} \mathcal{H}_{l+1} = n_u(l+1) + n_z = 10$
- minimum number of T : $T_{\min} = (n_u+1)(l+1) + n_z - 1 = 14$

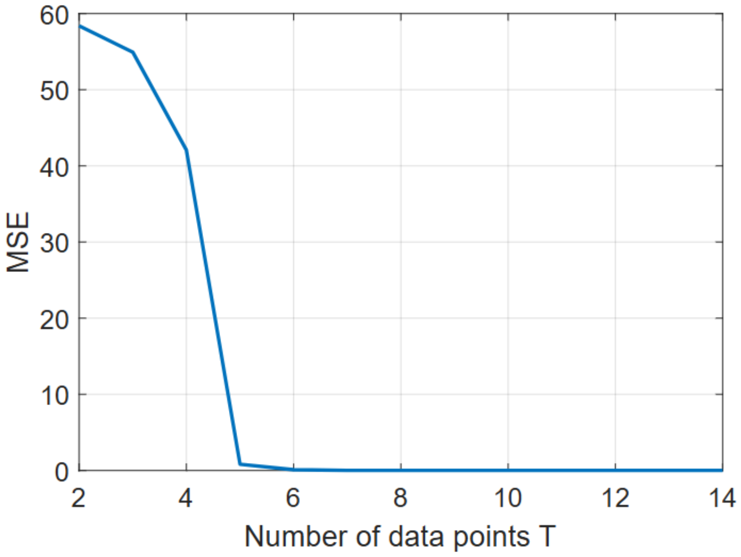


Fig. 1. Mean square error of the system's output signals with respect to the number of data points T.

$$\mathrm{MSE} = \frac{1}{N}\sum_{k=1}^{N} \|\mathbf{y}_k - \hat{\mathbf{y}}_k\|_2^2,$$

LQR Control Task

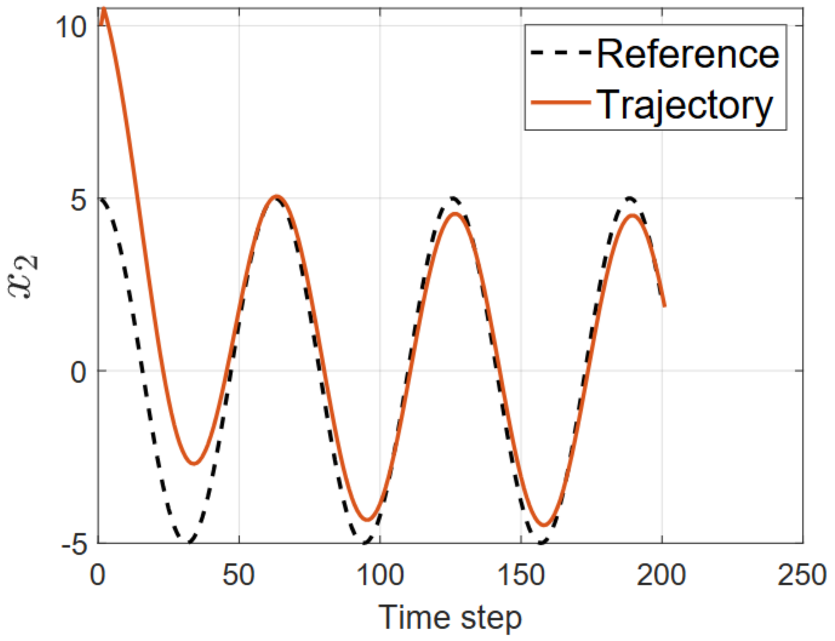


Fig. 2. Trajectory of x_2 over time steps under the LQR controller designed based on the identified model.

$$\hat{J} = \sum_{k=0}^{\infty} [\mathbf{z}_k^\top \hat{\mathbf{Q}} \mathbf{z}_k + u_k^\top \mathbf{R} u_k],$$

where

$$\hat{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{n_z \times n_z}, \quad \mathbf{Q} = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}, \quad \mathbf{R} = 0.01.$$

Feedback controller

$$u(k) = -K_{\mathrm{LQR}} \mathbf{z}_k$$

Conclusions

Conclusions

- We derive data collection conditions for nonlinear system identification, which admits Koopman linear embedding.
- We introduce an identifiability condition, which shows that the system is identifiable if the rank condition of a Hankel matrix of a certain depth is satisfied.
- We derive the minimum number of data points required for the nonlinear system identification with linear embedding.
- Future works : characterizing data conditions that ensure identifiability with no linear embedding.

Thank you!

Contact

Youngkyoung Kong
Department of Aerospace Engineering, Automation and Systems Research Institute
Seoul National University, Seoul, Korea
youngkyoung.kong@gmail.com