# Applications of System Identification

**Xiaoyu He,** Petr Ryšavý and Jakub Mareček | Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague, Czechia.
Mengjia Niu, Quan Zhou | Dyson School of Design Engineering, Faculty of Engineering, Imperial College London, UK.

## Abstract

Clustering of time series is a well-studied problem, with applications ranging from quantitative, personalized models of metabolism obtained from metabolite concentrations to state discrimination in quantum information theory. We consider a variant, where given a set of trajectories and a number of parts, we jointly partition the set of trajectories and learn linear dynamical system (LDS) models for each part, so as to minimize the maximum error across all the models. We present globally convergent methods and EM heuristics, accompanied by promising computational results. Additionally, we give an example of Krebs cycle, which can be observed as a biomedical application of systems identification (SI) in causal learning.

## Problem Preliminaries

- To search for the global optima of a problem, we developed relaxations to bound the optimal objective values in **non-convex Mixed-Integer Nonlinear Programs (MINLPs).**
- To extend the search to an operator in a Hilbert space with an unknown dimension, Zhou and Mareček applied **Non-Commutative Polynomial Optimization (NCPOP)** to SI.
- NCPOP is an operator-valued optimization problem defined on a **Hilbert space** with operators **X** and normalized state **ψ**. Given the input data p(x) and $q_i(X)$ , the standard form is

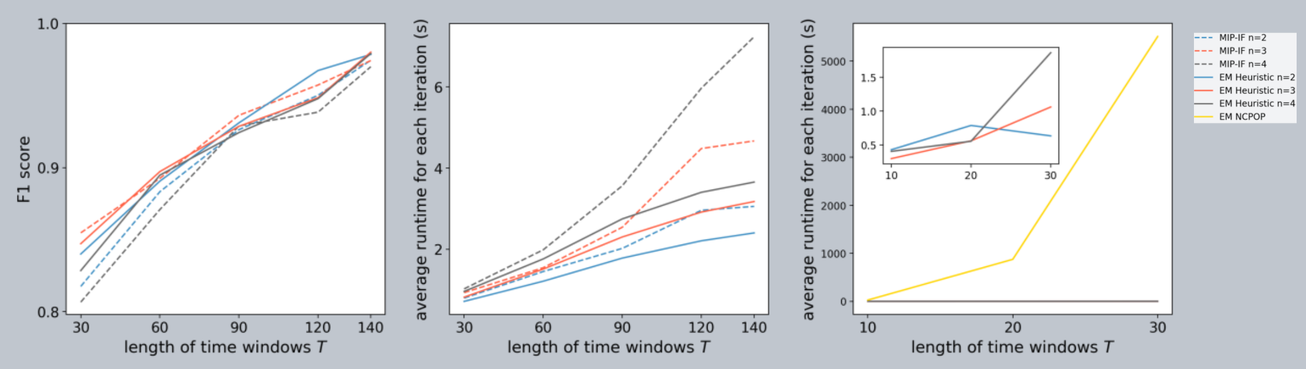
$$\begin{aligned} \underset{(\mathcal{H}, X, \psi)}{\text{minimize}} \quad & \langle \psi, p(X)\psi \rangle \\ \text{subject to} \quad & q_i(X) \succcurlyeq 0, i = 1, \ldots, m, \\ & \langle \psi, \psi \rangle = 1, \end{aligned}$$

where q_i(X) is positive semidefinite.

## EM Heuristic

- Finding an optimal trajectory for a set of trajectories is easier than clustering the trajectories. Therefore, as a complement to the NCPOP formulation, we provide an efficient Expectation-Maximization (EM) procedure for clustering time-series.

- **Experimental results:**

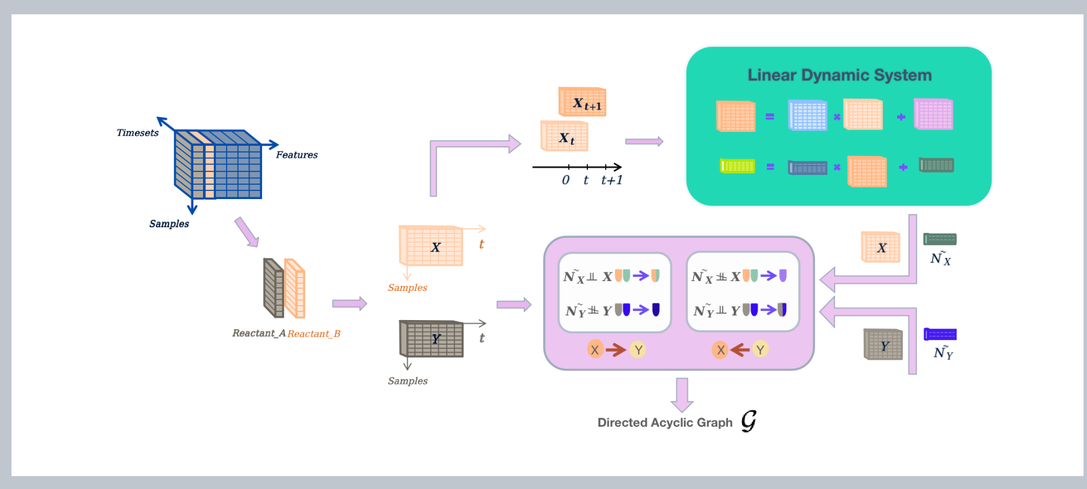



- As depicted in the above figures, our approach based on SI consistently outperforms conventional methods, particularly as the length of the time window increases.

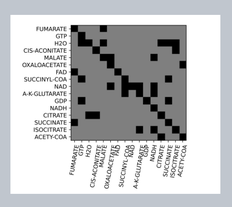
Full paper link

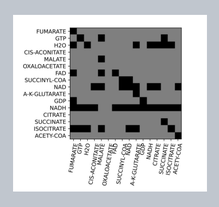
## Causal Learning & Future Work

- The learning of Nonlinear causal discovery with additive noise models (ANM) can be reduced to the learning of linear dynamic systems.




- We consider training a rich class of causal models from time-series data and we suggest the use of the Krebs cycle and models of metabolism more broadly. Here are adjacency matrices produced by the ground truth and prediction.




Full paper link

## References

- P. Belotti, C. Kirches, S. Leyffer, J. Linderoth, J. Luedtke, and A. Mahajan, "Mixed-integer nonlinear optimization," *Acta Numerica*, vol. 22, pp. 1–131, 2013.
- J. Peters, D. Janzing, and B. Schölkopf, "Elements of causal inference: foundations and learning algorithms." *The MIT Press*, 2017.
- Q. Zhou and J. Mareček, "Learning of linear dynamical systems as a non-commutative polynomial optimization problem," *IEEE Transactions on Automatic Control,* 2023.

**hexiaoyu@fel.cvut.cz** | The CoDiet project is funded by the European Union under Horizon Europe grant number 101084642 and supported by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 101084642].