METHODS AND APPROACHES TO BENCHMARK DATA-DRIVEN MODELING IN THE SPARSE AND NOISY DATA REGIMES

Alex Gorodetsky and Nicholas Galioto

[1]*University of Michigan*

ABSTRACT

Data-driven models have shown great success to learning dynamical systems with desired properties when data is plentiful and informative. However, many problems involved with dynamical systems have small data, both temporily and in the context of partial observability. Moreover, this data can often arise from noisy observations. In these regimes, the predominant approaches in the literature have largely focused on heuristic pre-processing techniques and additional regularization methods. In this talk we provide examples of how such heuristics can still struggle, even in simplified problems. We then describe how the source of these struggles is often incomplete modeling of the relationship between data-points of a time series. A large number of methods that rely on simple least-squares-based reconstruction of trajectories often implicitly presume that data in a time-series is independent and so a sum of squared errors represents an appropriate learning objective. We derive an alternative objective from standard Hidden Markov Modeling assumptions that shows such objectives are not appropriate, especially in the low-data regime where it is most important to extract all the relationships between the data. We then both optimize and sample the objective (in a Bayesian setting) and demonstrate improved reconstruction compared to existing approaches for both linear and nonlinear problems.

We demonstrate these facets on several problems that could serve as good benchmarks in the field. We first show that even a simple linear pendulum causes some traditional approaches to break down in our target data regime. We then show how canonical system-ID problems, such as the Wiener-Hammerstein problem can be modified to serve as appropriate benchmarks. Finally, we demonstrate a few problems where physics-constraints, such as conservation of energy in Hamiltonian dynamics, are important to incorporate into the learning approach.

REFERENCES

[1] Galioto, N., and Gorodetsky, A. A., "Bayesian system ID: optimal management of parameter, model, and measurement uncertainty." *Nonlinear Dynamics* 102.1 (2020): 241-267.
[2] Galioto, N., and Gorodetsky, A. A.. "Bayesian identification of Hamiltonian dynamics from symplectic data." *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020.
[3] Galioto, N, and Gorodetsky, A. A.. "A New Objective for Identification of Partially Observed Linear Time-Invariant Dynamical Systems from Input-Output Data." *Learning for Dynamics and Control*. PMLR, 2021.
[4] Galioto, N., and Gorodetsky, A. A., "Robust identification of non-autonomous dynamical systems using stochastic dynamics models." *arXiv preprint arXiv:2212.13902* (2022).