CPS:Medium:Safe Learning-Enabled Cyberphysical Systems, CNS-2038493

Co-PIs: M. Sznaier and O. Camps,

Robust Systems Lab Northeastern University


A. Comas


J. Miller


J. Zheng


O. Camps

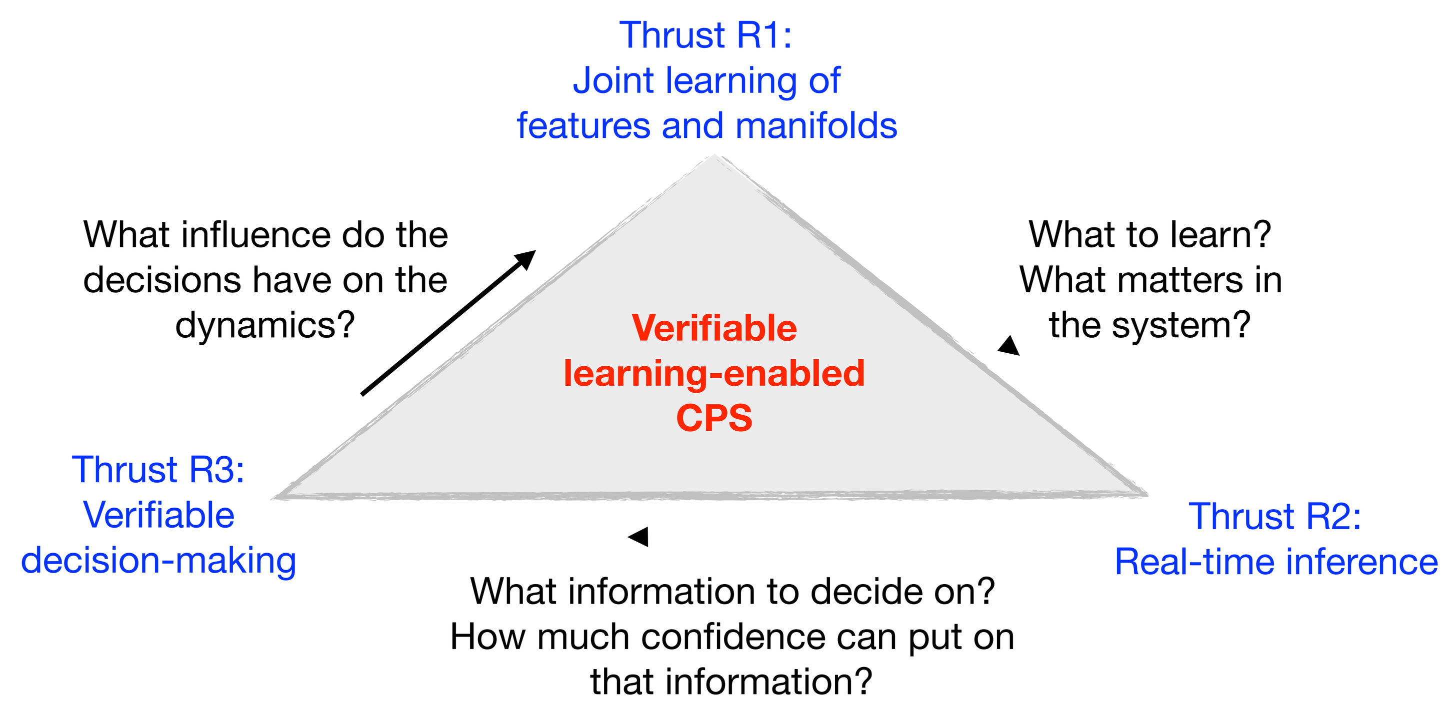

M. Sznaier

Motivation

Design autonomous CPS capable of safely operating in and adapting to previously unseen scenarios. (Humans can do it!)

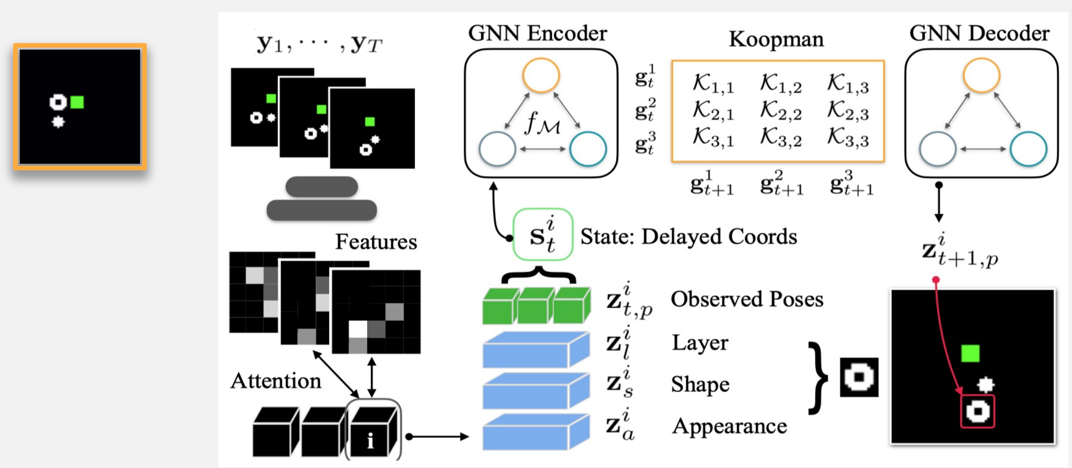
Challenges

- Lack of training data (often single execution).
- Need to act while learning (no re-do!).
- Actionable information sparsely encoded in large data sets.



Koopman-based Identification of Interacting Dynamics

- Goal: learn dynamics of multiple interacting agents.
- Main idea: Graph-based Object Centric Dynamic Mode Decomposition with Koopman Operators.
- Technical details:
  - Use delayed coordinates and attention-based tracking
  - Use a GNN to encode and decode the trajectories
  - Enforce linear manifold dynamics, with interpretable representations



2024 SYSID

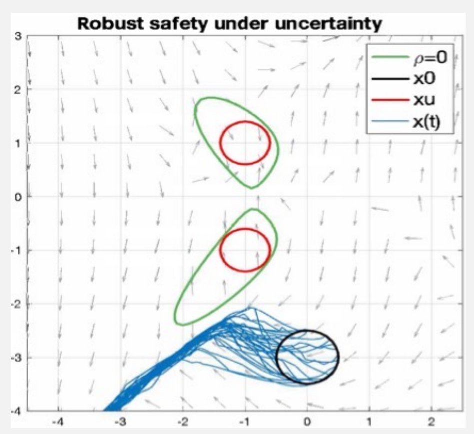
Verifiable Safe Data Driven Control

- Goal: data driven avoidance of an unsafe set.
- Main idea: enforce robust Nagumo-invariance .
- Technical details:
  - Prior: non linear dynamics of the form:
    $\dot{\mathbf{x}} = \mathbf{F}\phi(\mathbf{x}) + \mathbf{G}\gamma(\mathbf{x})\mathbf{u} + \eta \quad \mathbf{F}, \mathbf{G}$ unknown
  - Initial condition set: $\mathcal{X}_o$, bad set: $\mathcal{X}_u \doteq \{\mathbf{x} : h(\mathbf{x}) \geq 0\}$
  - Experimental data: $(\mathbf{u}(t_k), \mathbf{x}(t_k), \dot{\mathbf{x}}(t_k)),\ 0 \leq t_k \leq T$
  - Fact: bad set is avoided if there exist $\rho(x), \psi(x)$
    $\rho(\mathbf{x}) \geq 0,\ \forall \mathbf{x} \in \mathcal{X}_0,\ \rho(\mathbf{x}) < 0, \forall \mathbf{x} \in \mathcal{X}_u$ and

$$\nabla \cdot \left[\rho(\mathbf{x})\Big(\sum_i \alpha_i \mathbf{f}_i(\mathbf{x}(t)) + \mathbf{w}\Big) + \psi(\mathbf{x})\sum_j \beta_j \mathbf{g}_j(\mathbf{x})\right] - \rho(\mathbf{x})h(\mathbf{x}) > 0$$

for all plants that could have generated the data

  - Fact: reduces to a convex SDP via duality
  - Fact: only imposes $d\rho/dt < 0$ when $\rho = 0$



2024 IEEE LCSS, 2024 IEEE TAC

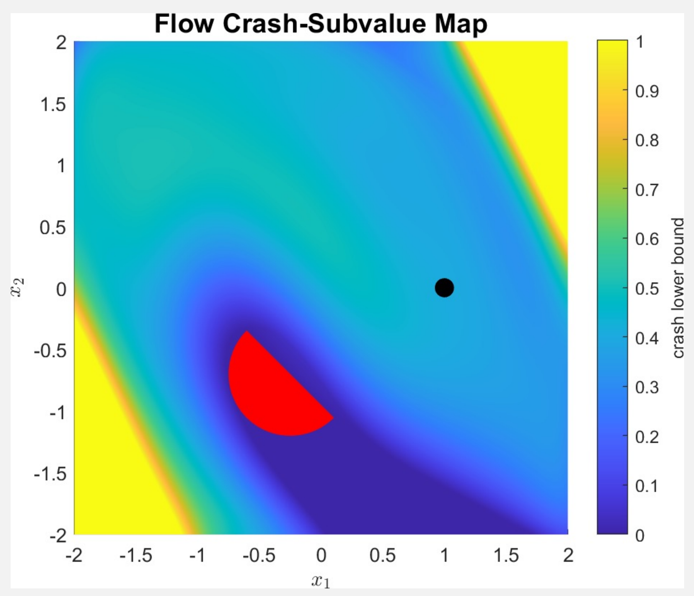
Safety Margins:

- Goal: Estimate how much uncertainty can be tolerated.
- Technical details:
  - Prior: non linear dynamics of the form:

$$\dot{x}(t) = f(t, x(t), w(t))$$
$$w(\cdot) \in \mathcal{W},\ t \in [0, T],\ x_0 \in X_0.$$

  - A reachability problem:

$$\begin{aligned} Q_z^* = \inf_{t,\ x_0,\ z,\ w}\ & z \\ & \dot{x}(t') = f(t', x(t'), w(t')), \dot{z}(t')) = 0, \quad \forall t' \in [0, T] \\ & J(w(t')) \leq z, \quad \forall t' \in [0, T] \\ & x(t \mid x_0, w(\cdot)) \in X_u \\ & w(\cdot) \in W,\ t \in [0, T],\ x_0 \in X_0, z \in [0, J_{\max}]. \end{aligned}$$



  - Reduces to convex optimization via occupation measures and SoS
  - Work in progress: extend to the data driven case

To appear, SIAM Sicon 2025

Application

- Public space monitoring to mitigate unsafe situations.


Monitoring social distancing

Active shooter scenario (joint with M. Siami)

Scientific Impact

- Rapprochement of Systems Theory, ML, Viability.
- Efficient extraction of actionable information from large data sets.
- Frugal, explainable architectures for dynamics oriented learning.

Broader Impact

- Certified safe learning enabled systems that can operate in close proximity to humans.
- Applications: safe autonomy, disaster mitigation, public space monitoring.
- Internships at Adobe, Mathworks, Google.

Award CNS-2038493