Verification of Neural-Network Control Systems by Integrating Taylor Models and Zonotopes

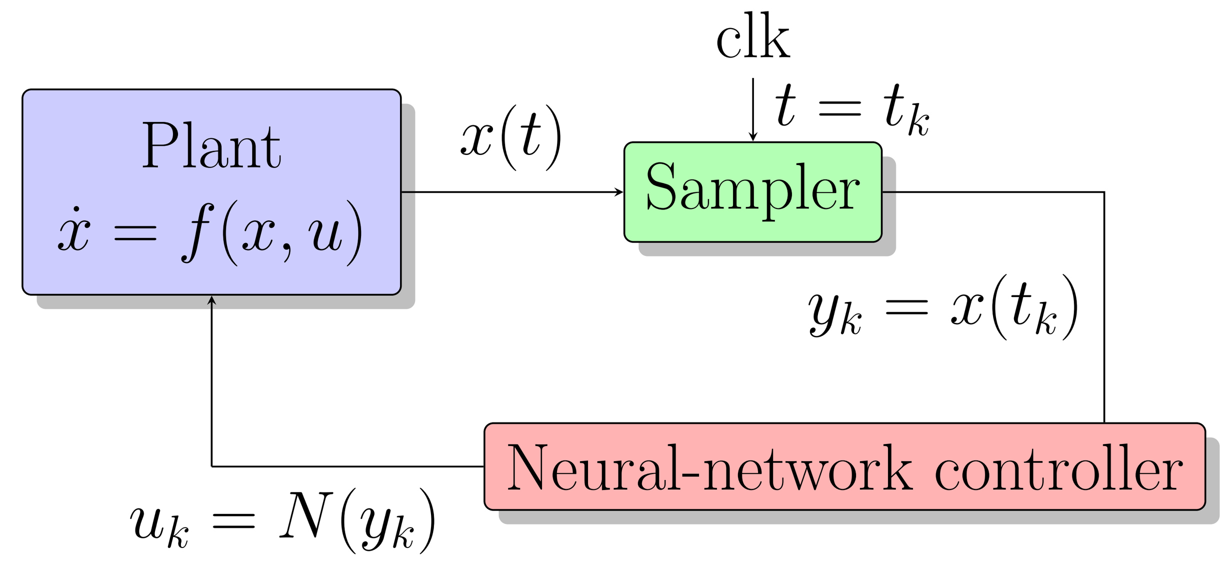
Paper

Code

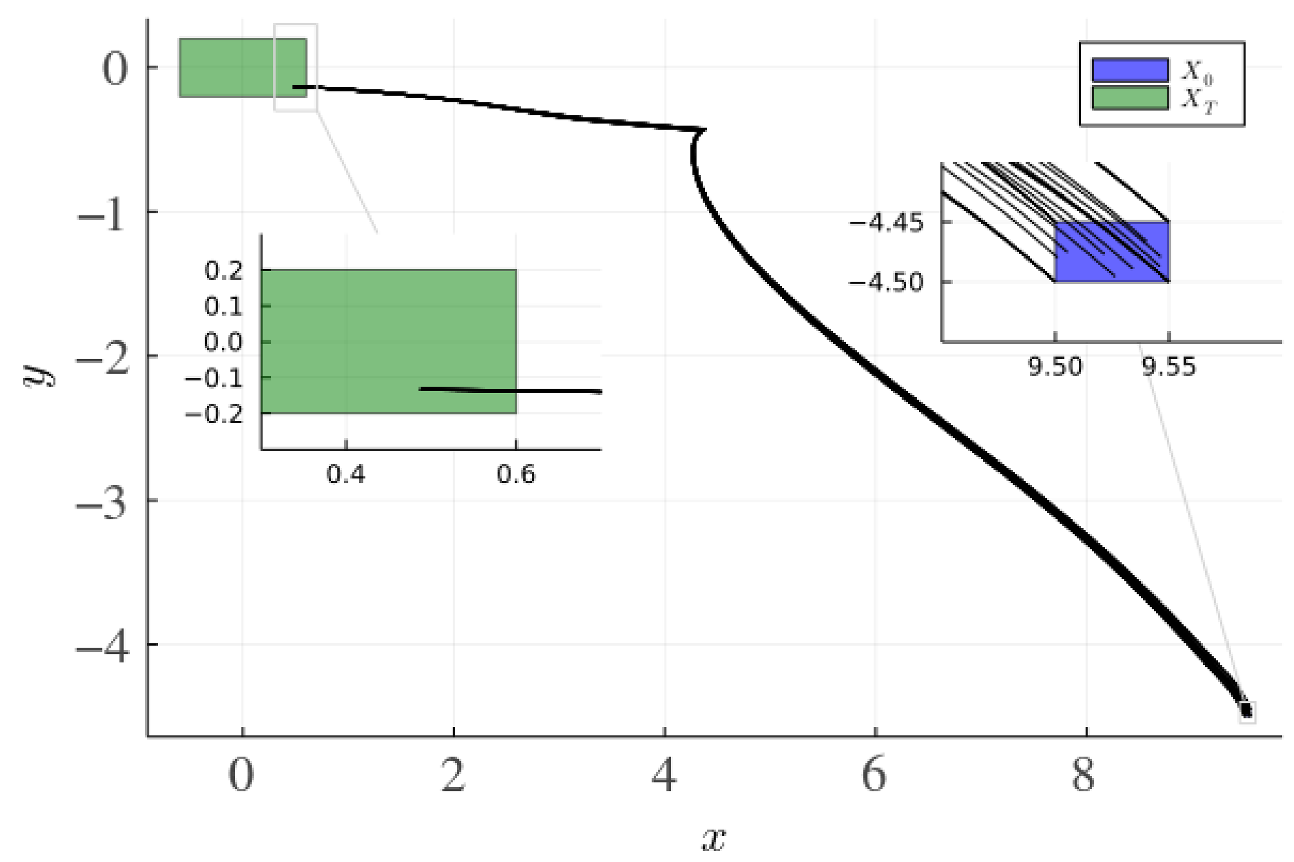
Christian Schilling[1], Marcelo Forets[2], Sebastián Guadalupe[2]

[1]Aalborg University, Denmark [2]Universidad de la República, Uruguay

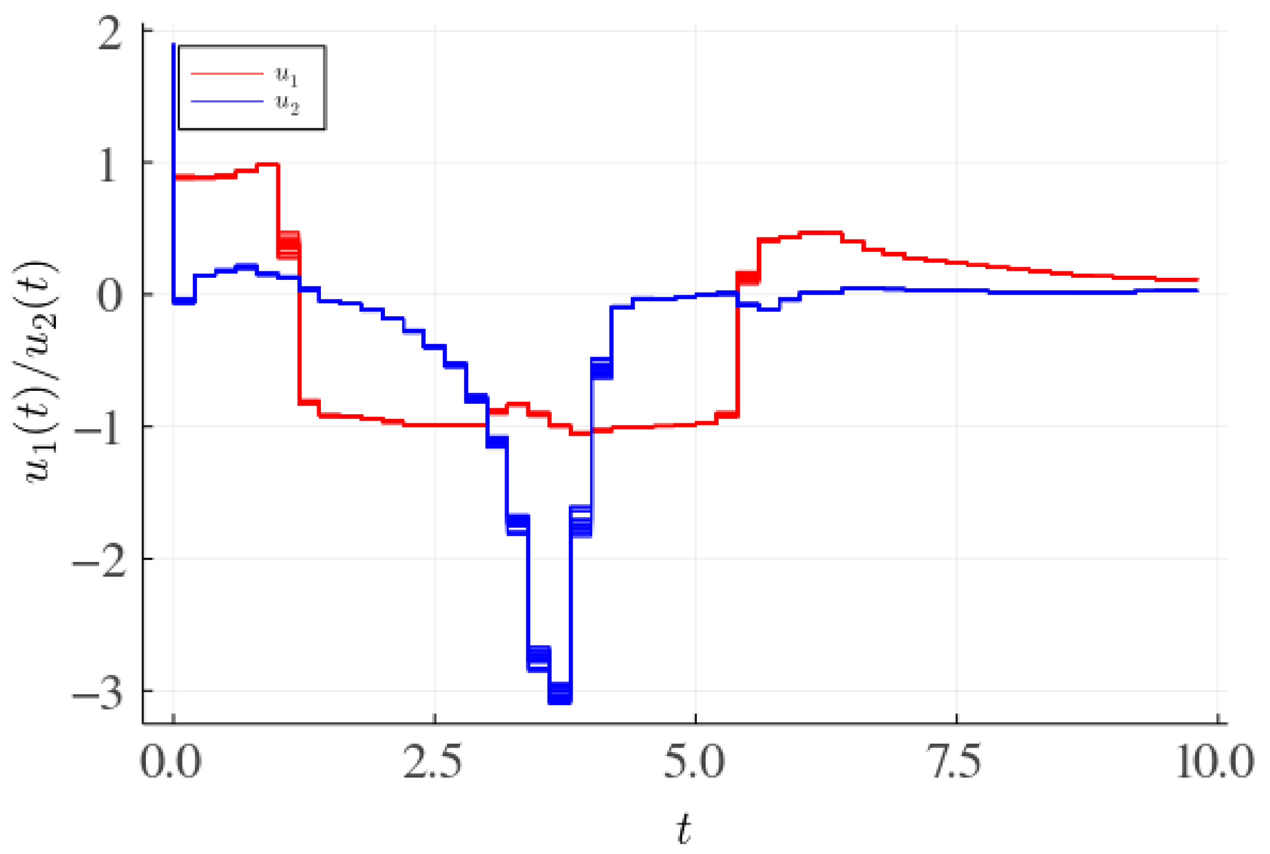
Neural-network control system



Example model: Unicycle



42 simulations from different initial conditions.

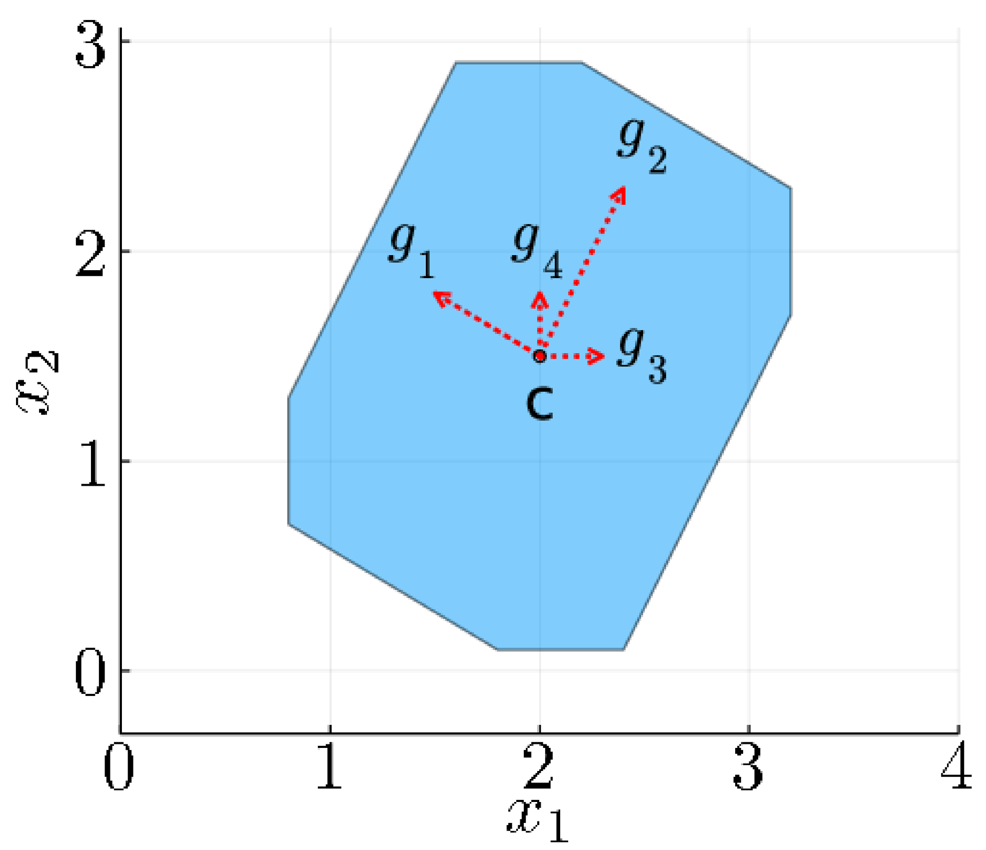


42 corresponding control signals.

Structured zonotope (SZ)

Characterized by center $c \in \mathbb{R}^n$ and generators $g_1, \dots, g_{2n} \in \mathbb{R}^n$ with $g_{n+1}, \dots, g_{2n}$ axis aligned, it defines the set of points

$$\left\{ c + \sum_j \zeta_j g_j \;\middle|\; \zeta_j \in [-1, 1] \right\}.$$



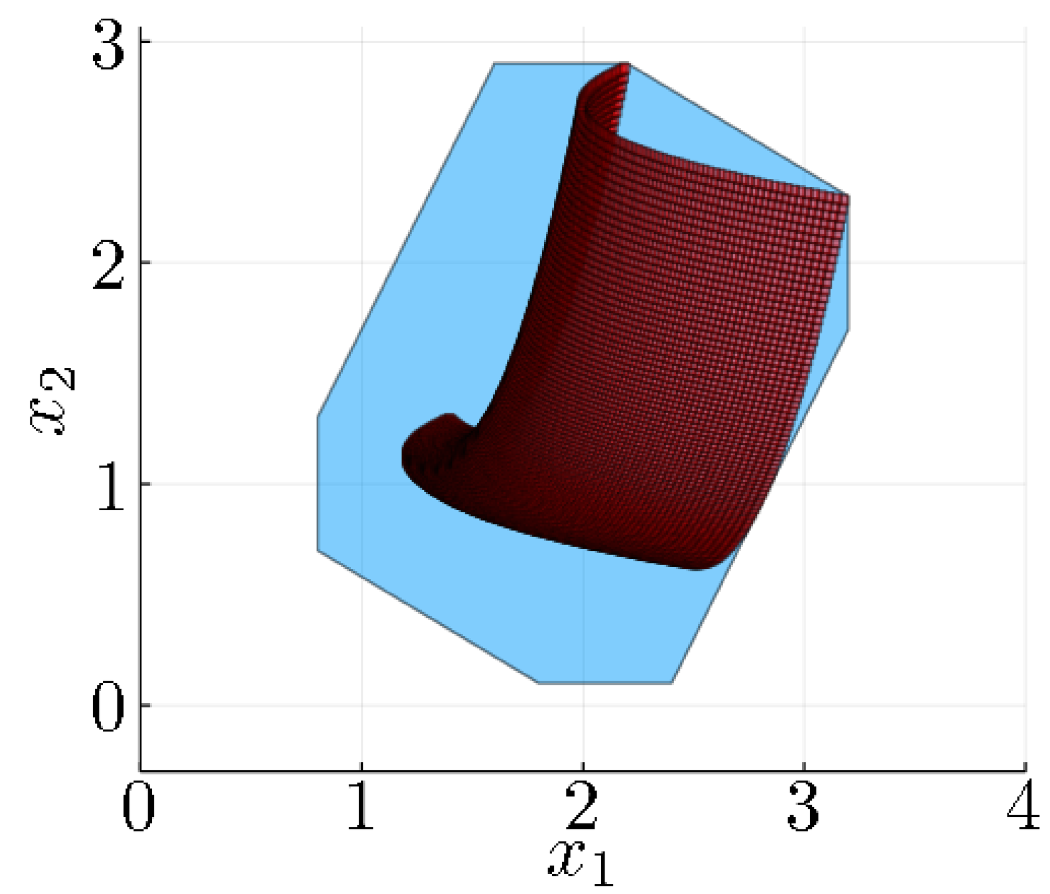
Challenge & approach

- Compute set of reachable states
- Set propagation through plant (TMRS) and neural network (SZ)
- Conversion between sets is exact for SZ→TMRS but only approximate for TMRS→SZ
- Conversion error accumulates
- Instead construct new TMRS from previous one and integrate SZ

- Implemented in JuliaReach
- First tool that solved all ARCH-COMP benchmarks

Taylor model

Characterized by vector of multivariate polynomials $p = (p_1, \dots, p_n)^T$, remainder $\Delta \subseteq \mathbb{R}^n$, and domain $\mathcal{D} \subseteq \mathbb{R}^n$, it defines an interval tube as the vector-valued function $p(x) + \Delta$.



Enclosure of a Taylor model in a structured zonotope.

Taylor-model reach set (TMRS)

An n-dimensional vector of Taylor models in one variable (time) with shared domain whose coefficients are multivariate n-dimensional polynomials.

Evaluation at a time point yields an n-dimensional Taylor model.

Enclosure of the reachable states for the example model