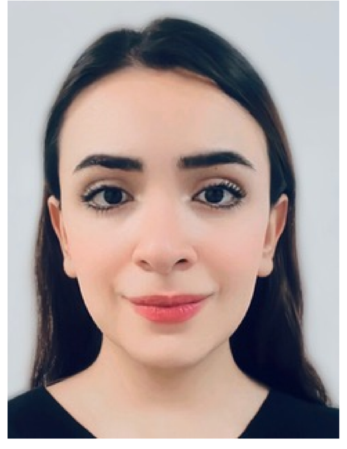
Data-Driven Safety Verification of Cyber-Physical Systems



Sara Taheri

sara.taheri@sosy.ifi.lmu.de

Supervisor: Majid Zamani & Matthias Althoff

CONVEY

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Motivation and Contribution

- Cyber-physical systems (CPS) become pervasive
- Many CPS are safety-critical, making it paramount to ensure their safe operation
- The majority of CPS are influenced by noise and uncertainty
- Models of CPS are either unknown or too complex to be of any use

CPS Models

A discrete-time stochastic control system (dt-SCS) is a tuple $\mathcal{S} = (X, U, V_m, w, f)$ where:

- $X \subseteq \mathbb{R}^n$ and $U \subseteq \mathbb{R}^m$ are the sets of state and input, respectively.
- w is a sequence of independent and identically distributed (i.i.d.) random variables on uncertainty space V_m.
- $f : X \times U \times V_m \to X$ is the state transition map such that:

$$x(t+1) = f(x(t), u(t), w(t)), \quad \forall t \in \mathbb{N}.$$

Safety Problem

Consider a dt-SCS $\mathcal{S}$, where the map f and the probability distribution of w are unknown. Consider a safety specification denoted by $\Psi = (X_0, X_u)$. System $\mathcal{S}$ is called safe with respect to Ψ, denoted by $\mathcal{S} \models \Psi$, if all trajectories of $\mathcal{S}$ started from the initial set $X_0 \subset X$ under a control policy C, never reach unsafe set $X_u \subset X$.

Safety Verification of dt-SCS

Definition 1: Control Barrier Certificate

Consider a dt-SCS $\mathcal{S}$ and a safety specification Ψ. Function $B : X \to \mathbb{R}_0^+$ is called a control barrier certificate (CBC) for $\mathcal{S}$ if there are constants $0 < \gamma < \lambda$ and a feedback controller $C : X \to U$ such that:

$$\begin{aligned} &B(x) \leq \gamma, && \forall x \in X_0, && \text{(a)}\\ &B(x) \geq \lambda, && \forall x \in X_u, && \text{(b)}\\ &\mathbb{E}\big[B(f(x, C(x), w)) \mid x\big] \leq B(x), && \forall x \in X \backslash X_u. && \text{(c)} \end{aligned}$$

Theorem 1: Safety Probability

Let $\mathcal{S}$ be a given dt-SCS with a safety specification Ψ. Suppose there is a CBC B and its associated controller C for the system $\mathcal{S}$. Then, one gets $\mathbb{P}\{\mathcal{S}_C \models \Psi\} \geq 1 - \frac{\gamma}{\lambda}$, where $\mathcal{S}_C$ represents the dt-SCS $\mathcal{S}$ controlled by C.

Data-driven Synthesis of CBC

Finding a CBC B and its corresponding controller C for a dt-SCS $\mathcal{S}$ is not possible, since the map f and the probability distribution of w are unknown.

(1) Considering CBC B and Controller C as two separate neural networks, $\mathbf{N}_b : \mathbb{R}^n \to \mathbb{R}_0^+$ and $\mathbf{N}_c : \mathbb{R}^n \to \mathbb{R}^m$, respectively. Then, collection of sample pairs $(x_i, u_i), i \in \{1, \dots, N\}$, from the sets of state and input, and also defining the loss function:

$$L = \sum_{\ell=1}^{4} \sum_{i=1}^{N} \text{ReLU}(g_\ell(x_i)),$$

$$\begin{aligned} g_1(x_i) &= -\mathbf{N}_b(x_i) - \eta, && \forall x_i \in X\\ g_2(x_i) &= \mathbf{N}_b(x_i) - \gamma - \eta, && \forall x_i \in X_0\\ g_3(x_i) &= -\mathbf{N}_b(x_i) + \lambda - \eta, && \forall x_i \in X_u\\ g_4(x_i) &= \mathbb{E}\big[\mathbf{N}_b(f(x_i, \mathbf{N}_c(x_i), w) \mid x\big] - \mathbf{N}_b(x_i) - \eta, && \forall x_i \in X \backslash X_u \end{aligned}$$

(2) Replacing the expectation term in g_4 with its empirical mean by using i.i.d. samples $w_j, j \in \{1, \dots, \hat{N}\}$, for each pair of $(x_i, u_i), i \in \{1, \dots, N\}$. Hence:

$$\bar{g}_4(x_i) = \frac{1}{\hat{N}} \sum_{j=1}^{\hat{N}} \mathbf{N}_b(f(x_i, \mathbf{N}_c(x_i), w_j)) - \mathbf{N}_b(x_i) + \delta - \eta, \forall x_i \in X \backslash X_u$$

where η is a negative robustness parameter ensuring that conditions in (a)-(c) are strongly satisfied, $\delta > 0$ is defined for the empirical mean approximation, and $\mathbf{N}_c(x_i)$ is bounded within U.

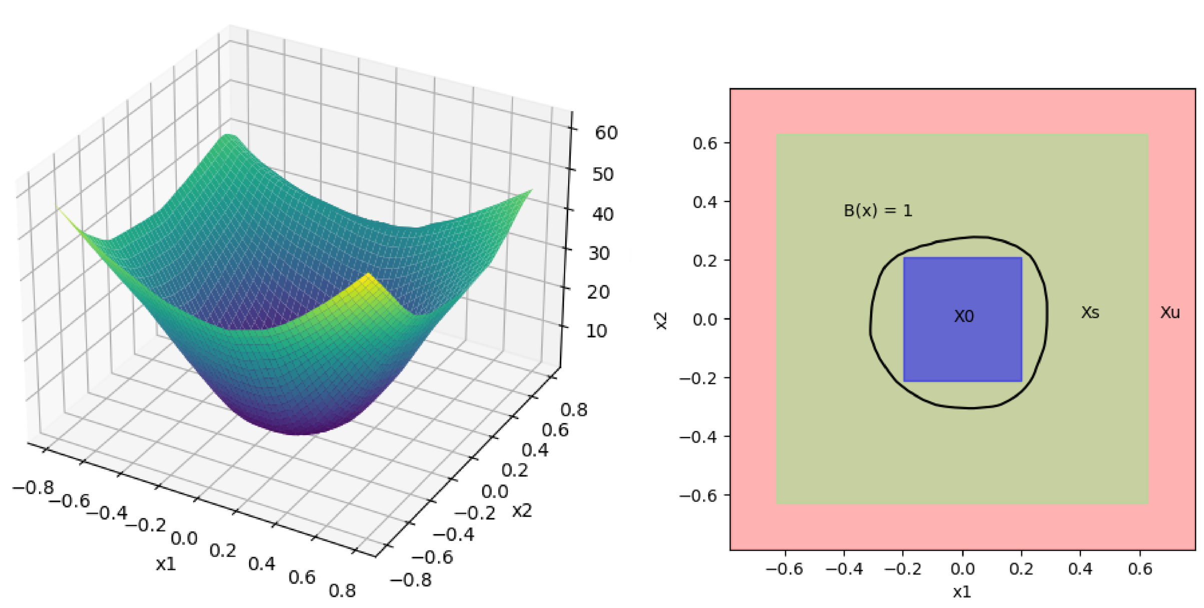
Correctness Guarantee of Neural Networks

Theorem 2: Correctness Guarantee

Consider a dt-SCS $\mathcal{S}$ and a safety specification $\Psi = (X_0, X_u)$. Assume that all constraints g_1, g_2, g_3, $\bar{g}_4$ are Lipschitz continuous with respect to pair (x, u), with a Lipschitz constant $\mathcal{L}$. Suppose $\hat{N} = \frac{\hat{M}}{\delta^2 \beta}$ for some $\delta > 0$ and $0 < \beta < 1$, where $\hat{M}$ is the upper bound for $\text{Var}(\mathbf{N}_b^*(f(x, \mathbf{N}_c^*(x), w))) \leq \hat{M}$ for trained neural networks $\mathbf{N}_b^*$ and $\mathbf{N}_c^*$ and for all $x \in X$. Collect N data pairs (x_i, u_i) with a quantization parameter ϵ. If $\mathcal{L}\epsilon + \eta \leq 0$, then $\mathbb{P}\left\{\mathcal{S}_{\mathbf{N}_c^*} \models \Psi\right\} \geq 1 - \frac{\gamma}{\lambda}$ with a confidence of at least $1 - \beta$.

Case Study

Consider a dt-SCS of an inverted pendulum with additive zero-mean Gaussian noise (standard deviation = 0.01). Assume $X = \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]^2$, $X_0 = \left[-\frac{\pi}{15}, \frac{\pi}{15}\right]^2$, $X \backslash X_u = \left[-\frac{\pi}{5}, \frac{\pi}{5}\right]^2$, and $U = [-10, 10]$. The parameters are set to $\beta = 0.001$, $\gamma = 1$, $\lambda = 25$, $\hat{N} = 100$, $\delta = 2$, and $\epsilon = 0.00157$. The neural network $\mathbf{N}_b$ comprises 100 neurons across each of the 5 hidden layers, while $\mathbf{N}_c$ consists of 25 neurons in each of its 3 hidden layers, with learning rates of $l_{r_b} = 10^{-4}$ and $l_{r_c} = 10^{-3}$, respectively. Then, we obtain $\mathbb{P}\{\mathcal{S}_{\mathbf{N}_c} \models \Psi\} \geq 0.96$ with a confidence of at least 99.76%.



The constructed CBC over X (left) and the γ-level of CBC (right).

References

[1] M. Anand et al. "Formally verified neural network control barrier certificates for unknown systems". In: *IFAC-PapersOnLine* 56.2 (2023), pp. 2431–2436.

[2] A. Salamati et al. "Data-driven verification and synthesis of stochastic systems via barrier certificates". In: *Automatica* 159 (2024), p. 111323.