Ensuring Safety Through Stochastic Control Barrier Functions
A Flexible Approach Based On Linear Bound Propagation

Tuin, H.; Hoek, L.; Mathiesen, Frederik Baymler

Publication date
2025

Document Version
Final published version

Published in
Book of Abstracts 44th Benelux Meeting on Systems and Control

44^{th} Benelux Meeting

on

Systems and Control

March 18 – 20, 2025

Egmond aan Zee, The Netherlands

Book of Abstracts

The 44^{th} Benelux Meeting on Systems and Control is sponsored by

disc dutch institute of systems and control

Raffaella Carloni, Javier Alonso-Mora, Janset Dasdemir, and Erjen Lefeber (Eds.)

Book of Abstracts - 44^{th} Benelux Meeting on Systems and Control

University of Groningen
PO Box 72
9700 AB Groningen
The Netherlands

ISBN (PDF without DRM): 978-94-034-3117-8

Ensuring Safety Through Stochastic Control Barrier Functions: A Flexible Approach Based On Linear Bound Propagation

Koen Tuin Lucas Hoek Frederik Baymler Mathiesen
Delft University of Technology
{h.tuin, l.hoek}@student.tudelft.nl, f.b.mathiesen@tudelft.nl

1 Introduction

Safety-critical systems are increasingly incorporating machine learning-based components [2]. This is problematic due to the lack of explainability and robustness against adversarial attacks, prohibiting safety guarantees. Recent efforts have employed Stochastic Control Barrier Functions (SCBFs) to bound the safety probability for stochastic systems, but this method relies on a limiting assumption of concavity of the barrier function [1]. We relax this assumption using Linear Bound Propagation (LBP).

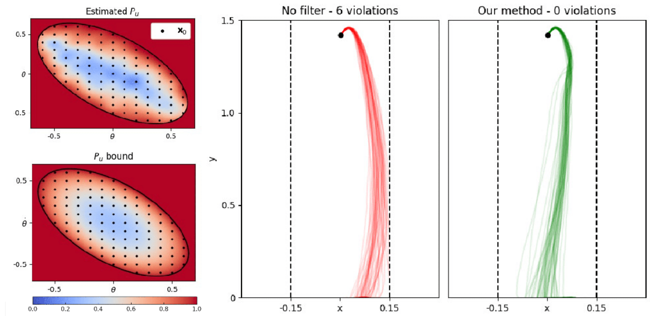

Figure 1: (**left**) Inverted pendulum over a time horizon of 1s. The bound on $P_u = 1 - P_s$ is computed using Eq. (2). (**right**) 50 trajectories of the lunar lander with and without safety filter. The nominal controller enters the unsafe set 6 times, while the safe controller does not leave the safe set.

2 Method

Consider a stochastic discrete-time system

$$\mathbf{x}(k+1) = f(\mathbf{x}(k), u(k)) + \mathbf{v}(k) \tag{1}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state, $u(k) \in U \subset \mathbb{R}^m$ is the control input, and $\mathbf{v}(k)$ is additive noise. Let $K_{nom} : \mathbb{R}^n \to \mathbb{R}^m$ be a given nominal controller. If the closed-loop system satisfies the SCBF condition [1], that is, there exists a function $h : \mathbb{R}^n \to \mathbb{R}$ with an upper bound $M \geq 0$ and constant $\alpha \in (0,1)$ such that the inequality below holds for all $x \in \mathbb{R}^n$

$$\mathbb{E}[h(f(x, K_{nom}(x)) + \mathbf{v})] \geq \alpha h(x), \tag{2}$$

then the system is guaranteed safe up to time step $K \in \mathbb{N}$ with probability $P_s \geq \frac{h(x(0))}{M} \alpha^K$. Furthermore, the SCBF condition can be used to design a safety filter, by solving the following stochastic Quadratic Programming (QP) problem at each time step for a given function h

$$K(x) = \underset{u \in U}{\text{argmin}} \quad \|u - K_{nom}(x)\|^2 \tag{3}$$

$$\text{s.t.} \quad \mathbb{E}[h(f(x,u) + \mathbf{v})] \geq \alpha h(x) \tag{4}$$

Solving Eq. (3)-(4) is hard: the dynamics may be unknown, evaluating the expectation may have no analytical solution, and the constraint may be non-concave in u. Despite these challenges, the QP problem must be solved in real-time due to the dependence on x, which is uncountable. Thus, we require sound approximations.

We identify the unknown dynamics of the system using a neural network (NN). Future work will focus on incorporating non-asymptotic bounds between NN and the underlying dynamics [3]. To handle the complexity of the NN and of h, we employ LBP from neural network verification [4]: computing linear functions in u that bound the output of the NN, given a compact input set U. The result is a linear lower bound $L_{LBP}(u) \leq \mathbb{E}[h(f(x,u) + \mathbf{v})]$ for all $u \in U$. Then replacing Eq. (4) with the following constraint results in an (approximately) minimally-invasive safe action:

$$L_{\text{LBP}}(u) \; \geq \; \alpha h(x). \tag{5}$$

This is a standard QP problem of m decision variables, hence easy to solve in real-time.

3 Results

We demonstrate our approach on two benchmarks. In the left of Fig. 1, we apply the SCBF to an *inverted pendulum*, using LBP. The results confirm that the empirical safety probability aligns with the derived theoretical bounds. The right of Fig. 1 shows the Gym Lunar Lander. Our filter successfully corrects the policy, ensuring safe trajectories with only minimal modifications to the nominal controller.

References

[1] Ryan K Cosner et al. “Robust Safety under Stochastic Uncertainty with Discrete-Time Control Barrier Functions”. In: *Robotics: Science and Systems*. 2023.

[2] Nicholas Roy et al. *From Machine Learning to Robotics: Challenges and Opportunities for Embodied Intelligence*. 2021. arXiv: 2110.15245 [cs.RO].

[3] Matthew Wicker et al. “Adversarial robustness certification for bayesian neural networks”. In: *International Symposium on Formal Methods*. 2024.

[4] H. Zhang et al. “Efficient neural network robustness certification with general activation functions”. In: *Advances in Neural Information Processing Systems*. 2018.