Safety Certification for Learning-Based Control

Technical University of Munich

TUM School of Computation, Information and Technology

Chair of Robotics, Artificial Intelligence and Real-time Systems

Background

Superior control performance is typically achieved by using sophisticated control algorithms, which require fine-tuned parameters. Also, learning-based control schemes have recently attracted a lot of interest and have been successfully applied to robotic systems [3]. However, in applications such as autonomous driving or human-robot collaboration, ensuring safety is paramount when applying these performance-oriented controllers.

Due to the complexity of these performance-oriented controllers, verifying safety is usually very challenging. Combining a safe backup controller with a given performance-oriented controller is a powerful yet simple approach to tackle this challenge: If required, the safe backup controller is used to modify desired performance-oriented control inputs in a minimally invasive way. Thereby, safety can be ensured without overly compromising performance. This approach has been applied in various domains such as safe reinforcement learning [10, 2], human-in-the-loop control [7] or motion planning for autonomous vehicles [8].

Description

In our previous work [9], we proposed an algorithm for filtering potentially unsafe inputs for nonlinear systems. At the core of this algorithm is the computation of a set of initial states that can be safely steered into a given target set. If the control input proposed by the performance-oriented controller keeps the system within reach of the target set, the proposed control input is applied. Otherwise, we override this control input by a safety-preserving control input.

However, the computation of the initial set in [9] relies on linearizing the nonlinear system and linear control laws which can lead to conservative results [4] that might overly compromise the resulting control performance. Therefore, the goal of this thesis is the extension of the algorithm in [9] to polynomial controllers [4].

All programming will be done in MATLAB and the final implementation should be integrated into our toolbox AROC [6].

This thesis provides you with the opportunity to get to know an advanced control algorithm that provides formal guarantees. Moreover, you gain/deepen your knowledge in control theory, numerical optimization, and reachability analysis, which is a tool for formal verification.

Tasks

- Familiarization with our open-source MATLAB tool CORA [1].
- Familiarization with the control algorithms in [9, 5, 4].
- Extension of the linearization-based approach in [9] to polynomial controllers [4].
- Extension of your algorithm (from the previous step) to optimizing the initial set and the safety-preserving controller.
- Extension of your algorithm to compute non-convex initial sets.
- Comparison of the implemented algorithms using different benchmark systems from our toolbox AROC [6].
- *Optional:* Comparison of the implemented algorithms with a state-of-the-art approach, such as [11].
- Documentation of your results.

Supervisor:
Prof. Dr.-Ing. Matthias Althoff

Advisor:
Lukas Schäfer, M.Sc.

Research project:
justITSELF

Type:
Master's thesis

Research area:
Safety-preserving control, robust control, formal methods

Programming language:
MATLAB

Required skills:
Background in control theory and dynamical systems. Background in numerical optimization beneficial.

Language:
English, German

Date of submission:
2. Juni 2022

For more information please contact us:

E-Mail: lukas.schaefer@tum.de

Internet: www6.in.tum.de

References

[1] Matthias Althoff. An introduction to cora 2015. In *ARCH@ CPSWeek*, pages 120–151, 2015.

[2] Osbert Bastani. Safe reinforcement learning with nonlinear dynamics via model predictive shielding. In *2021 American Control Conference (ACC)*, pages 3488–3494, 2021.

[3] Jaime F. Fisac, Anayo K. Akametalu, Melanie N. Zeilinger, Shahab Kaynama, Jeremy Gillula, and Claire J. Tomlin. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Transactions on Automatic Control*, 64(7):2737–2752, 2019.

[4] Victor Gaßmann and Matthias Althoff. Verified polynomial controller synthesis for disturbed nonlinear systems. *IFAC-PapersOnLine*, 54(5):85–90, 2021. 7th IFAC Conference on Analysis and Design of Hybrid Systems ADHS 2021.

[5] Felix Gruber and Matthias Althoff. Computing safe sets of linear sampled-data systems. *IEEE Control Systems Letters*, 5(2):385–390, 2021.

[6] Niklas Kochdumper, Felix Gruber, Bastian Schürmann, Victor Gaßmann, Moritz Klischat, and Matthias Althoff. Aroc: A toolbox for automated reachset optimal controller synthesis. In *Proceedings of the 24th International Conference on Hybrid Systems: Computation and Control*, HSCC '21, New York, NY, USA, 2021. Association for Computing Machinery.

[7] Ian M. Mitchell, Jeffrey Yeh, Forrest J. Laine, and Claire J. Tomlin. Ensuring safety for sampled data systems: An efficient algorithm for filtering potentially unsafe input signals. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7431–7438, 2016.

[8] Christian Pek, Stefanie Manzinger, Markus Koschi, and Matthias Althoff. Using online verification to prevent autonomous vehicles from causing accidents. *Nature Machine Intelligence*, 2(9):518 – 528, 2020.

[9] Bastian Schürmann, Moritz Klischat, Niklas Kochdumper, and Matthias Althoff. Formal safety net control using backward reachability analysis. *IEEE Transactions on Automatic Control*, pages 1–1, 2021.

[10] Kim P. Wabersich and Melanie N. Zeilinger. Linear model predictive safety certification for learning-based control. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 7130–7135, 2018.

[11] He Yin, Murat Arcak, Andrew Packard, and Peter Seiler. Backward reachability for polynomial systems on a finite horizon. *IEEE Transactions on Automatic Control*, 66(12):6025–6032, 2021.

Technical University of Munich

TUM School of Computation, Information and Technology

Chair of Robotics, Artificial Intelligence and Real-time Systems