Reachability Analysis Based Safety Falsification

Background

While formal verification aims to prove correct functionality of safety critical systems, safety falsification methods explicitly try to determine system trajectories that violate given safety specifications. The determined falsifying trajectories do not only prove that the system is not safe, but also provide the system designer with helpful insights about the system behavior.

Description

In order to determine falsifying trajectories, safety falsification methods try to find a suitable initial state and a suitable input signal. A recently developed concept enables to directly extract the initial state for the falsifying trajectory from a computed reachable set. Since this reduces the search space for the falsification problem to finding a suitable input signal, this novel concept potentially results in significant speed-ups compared to current state-of-the-art falsification approaches. The goal of thesis is to combine the novel concept for finding the initial state with different approaches for determining a suitable input signal.

The thesis includes the following steps: First, a literature review about safety falsification is performed to determine different methods for finding suitable input signals and identify the current state-of-the-art in the field of research. Next, the new falsification approach resulting from the combination of the new reachability analysis based method to determine the initial state with a suitable method for finding the input signal should be implemented using the MATLAB based reachability analysis toolbox CORA [1]. Finally, the performance of the novel approach is evaluated in comparison with existing state-of-the-art approached for safety falsification.

Falsification problem: determine a trajectory starting in the initial set (orange, green) which violates a given specification (red).
Tasks

- Literature review about safety falsification
- Implementation of the novel approach using CORA
- Evaluation of the performance of the novel approach compared to existing state-of-the-art algorithms

References