KTH ROYAL INSTITUTE
OF TECHNOLOGY

Safe Learning for Control

Caroline Heidenreich

May 26, 2017

Motivation

Why do we want to do safe learning in control?

- Model-based control suffers often from poor model accuracy
  $\Rightarrow$ Learning Control
- But: Reinforcement learning algorithms not designed for satisfying constraints so we need an additional safety-preserving controller
  $\Rightarrow$ Safe Learning Control

Example

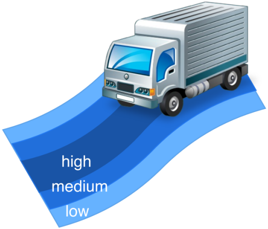
- Autonomous vehicle with partly known model
- Task: learn control with reinforcement learning without driving off the road
- To simplify, we only look at the truck's position

Algorithm

Markov Decision Process

- Discretise states and actions.
- Assign rewards to each state-action pair.
- Decide for an objective function that the agent should maximise.

Reward function



Objective function

$$R_t = \sum_{k=0}^{T} \gamma^k r_{t+k+1}$$

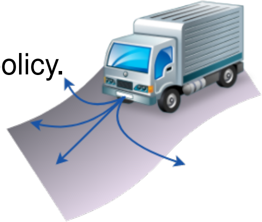
Algorithm

Reinforcement Learning

General Idea: Finding the optimal policy by playing actions and receiving rewards

1. The agent chooses an action.
2. The agent receives an external reward.
3. The agent updates its policy dependent on the received reward.

✓ The policy will converge to the optimal policy.
✗ No safety guarantees.



Algorithm

Safe Set Calculation

How can we ensure safety with uncertain dynamics?

- We treat the unmodelled dynamics as a bounded disturbance.
- With Hamilton-Jacobi-Isaacs (HJI) reachability analysis, we can determine for each state if it is safe or not.

Algorithm

Safe Set Calculation

- At the borders of the safe set: Apply safe control.
- Within the safe set: Reinforcement learning.

✓ Now we can learn a control without ever leaving the road.
✗ The safe set might be very small due to a conservative initial disturbance estimate.

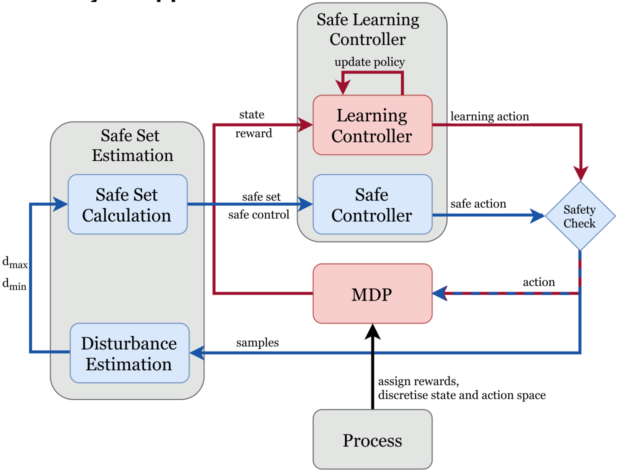
Algorithm

Disturbance Estimation with Gaussian Processes

- Update the conservative initial disturbance range with Gaussian Process (GP) regression in the light of data.
- GP regression: Non-parametric regression method that gives:
  a. an estimate for the disturbance.
  b. a measure how certain this estimate is.

Algorithm

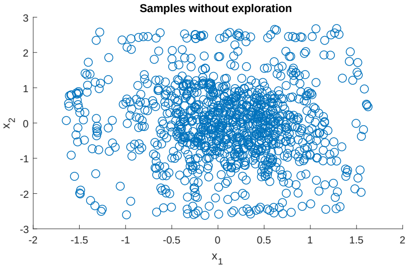
Summary of Approach

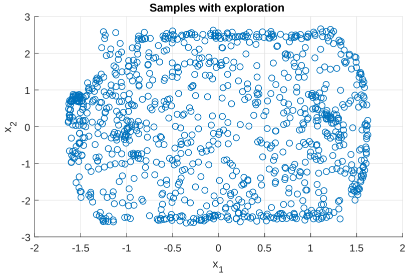


Algorithm

Exploration

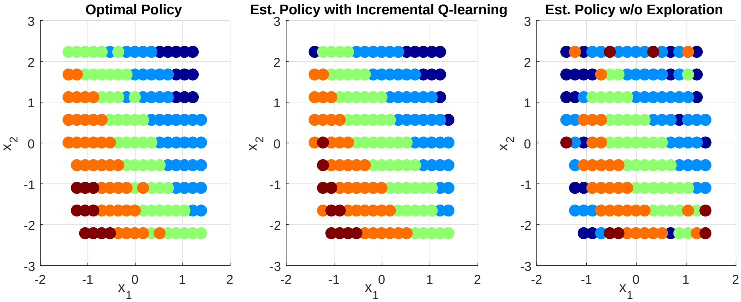
- Trade-off between exploration and exploitation.
- Our algorithm incorporates exploration explicitly by employing **Incremental Q-learning**





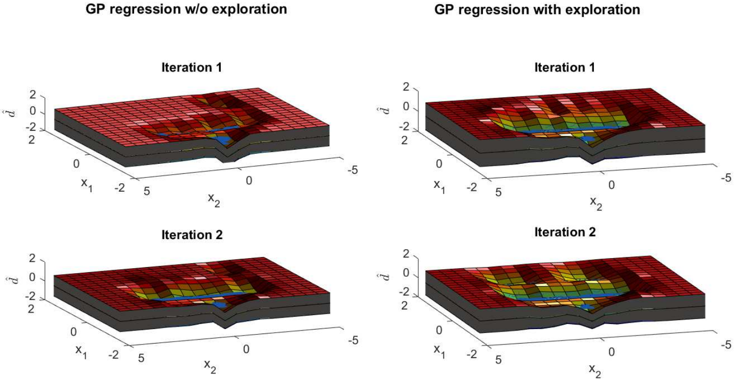
Experimental Results

Policy Estimation



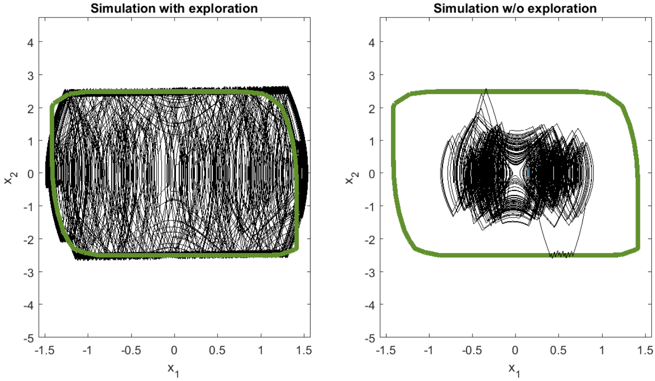
Experimental Results

Disturbance Estimation



Experimental Results

Simulation



Conclusions and Future Work

- Promising approach.
- Better results by incorporating exploration.

Some work remains to be done:

- Joint design of safety and learning loop.
- Recursive estimation of disturbance bounds.
- Formal guarantees for the whole algorithm.

difficult