Principles of Safe Autonomy
ECE 484 Lecture 3: Safety and invariance

Professors: Sayan Mitra

Tianchen Ji (tj12) Pranav Sriram (psriram2)

Haoyuan You (hy19) Ninghan Zhong (innghan2)

Last time: Automata→ invariance

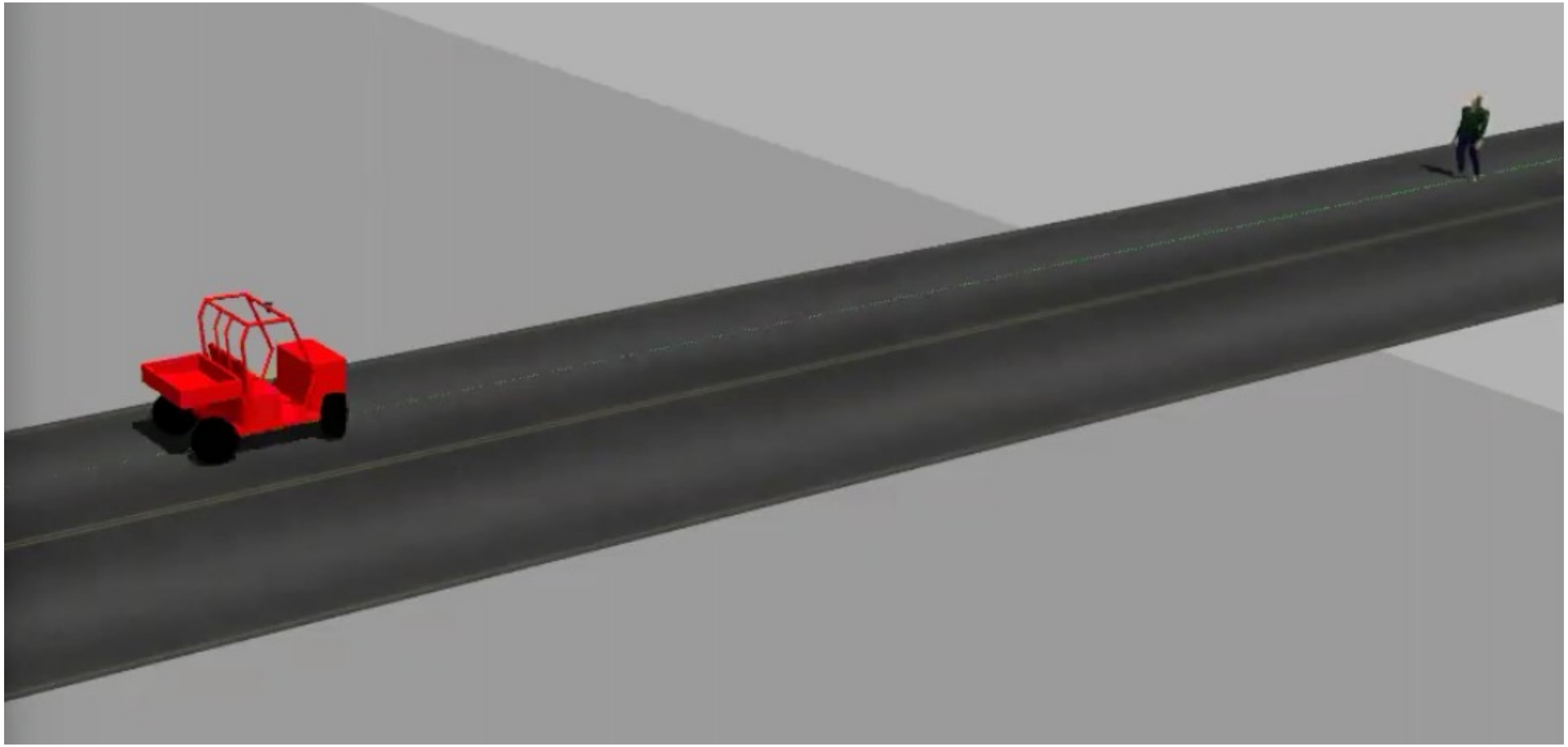
- Automaton: $A = \langle Q, Q_0, D\rangle$; nondeterminism $D \subseteq Q \times Q$
  - For any state $q \in Q, D(q) \subseteq Q$
  - For any set of states $S \subset Q, Post(S) := \cup_{q \in S} D(q)$
- Executions: $\alpha = q_0 q_1 \ldots q_k$
- Safety requirement $Unsafe \subseteq Q$
- Testing: Does there exist and execution $\alpha = q_0 \ldots q_k$ such that $q_k \in Unsafe$?
- Safety proof *or verification:* Show that there is no such execution
  1. One possible way: $\cup_{k=0}^{\infty} Post^k(Q_0) \cap Unsafe = \emptyset$ --- generally hard
  2. Invariance trick: Find $I \subseteq Q$ such that (i) $Q_0 \subseteq I$ and (ii) $Post(I) \subseteq I$ then $Post^k(Q_0) \subseteq I$ [Proposition 2]

  This is nice because then instead of 1. we can check $I \cap Unsafe = \emptyset$

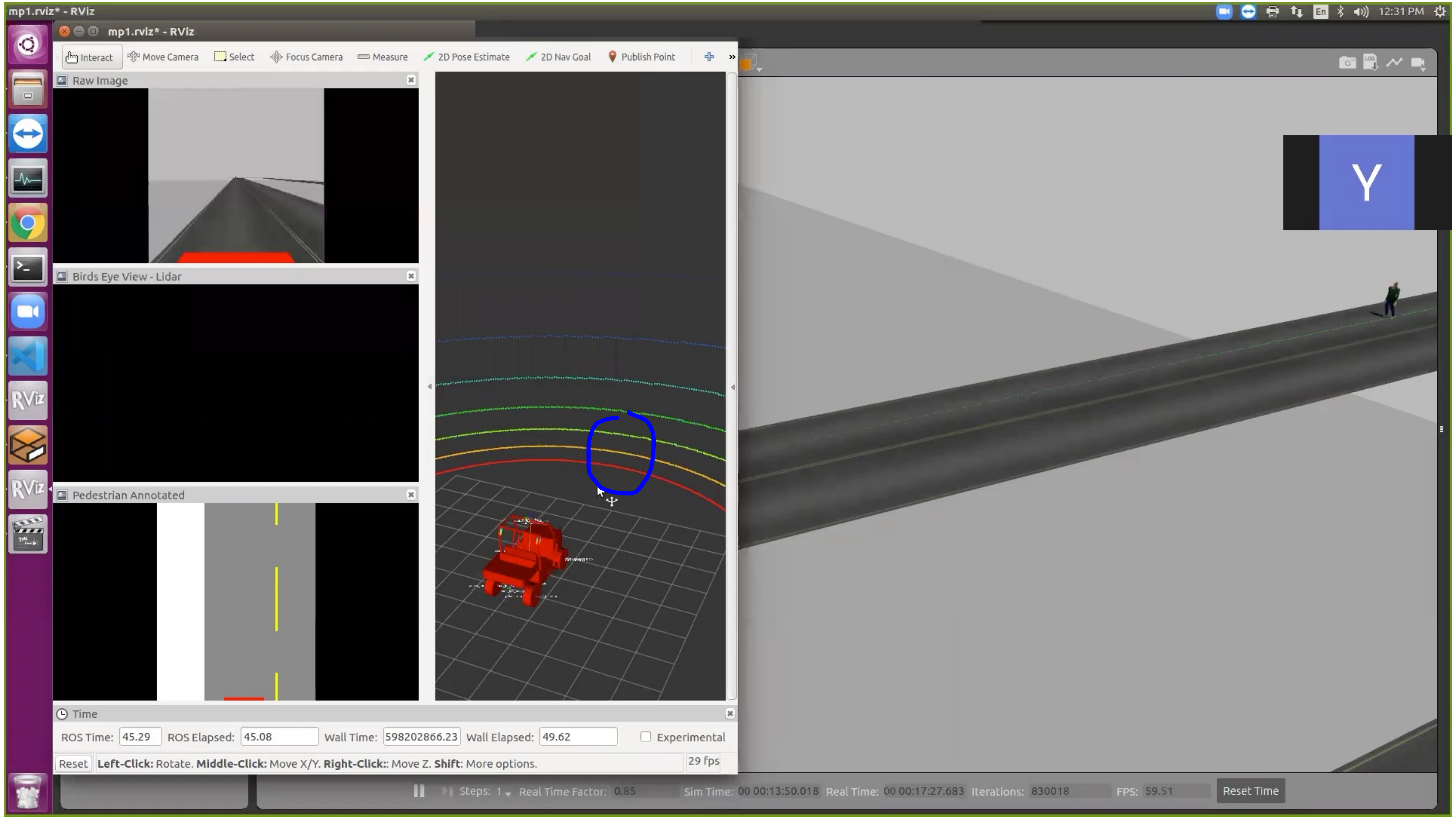
Roadmap

- Prove Proposition 2
- Guess I for AEB example and check it with Prop 2
- Discuss limits and consequences of trick

Model (switch to notes)



MP0: Simulate model for testing



“All models are wrong, some are useful.”

Wrong and useless

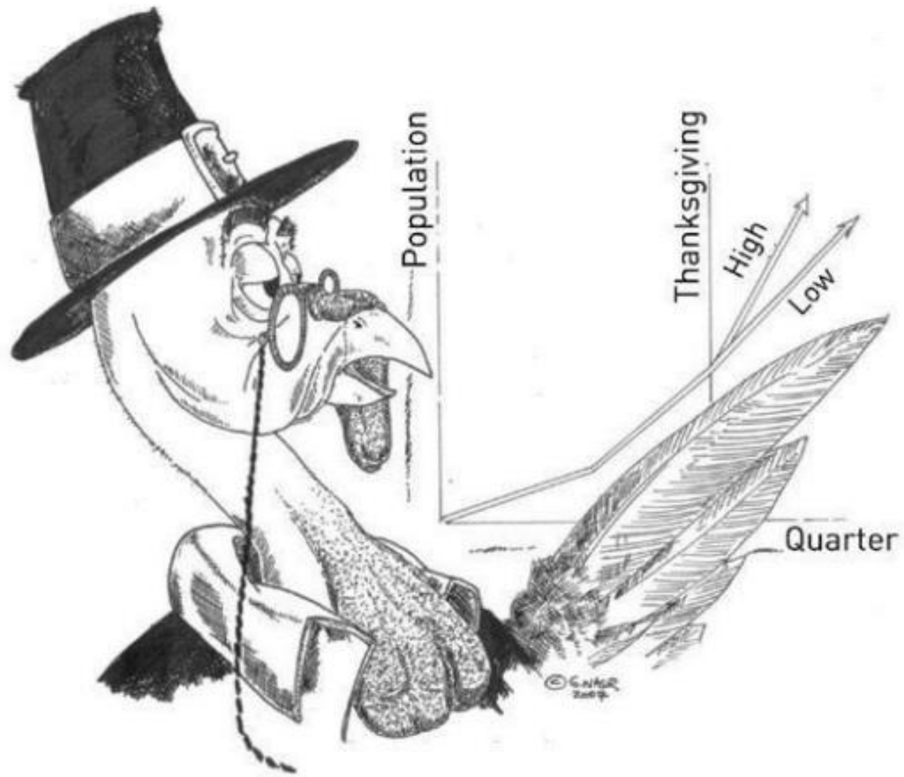
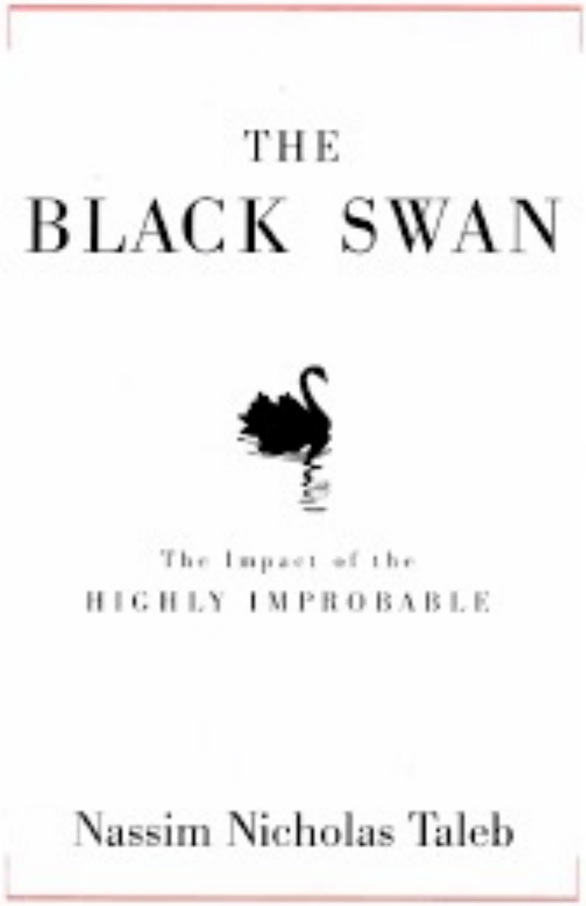
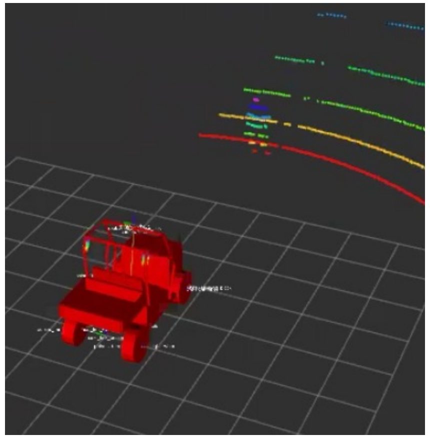


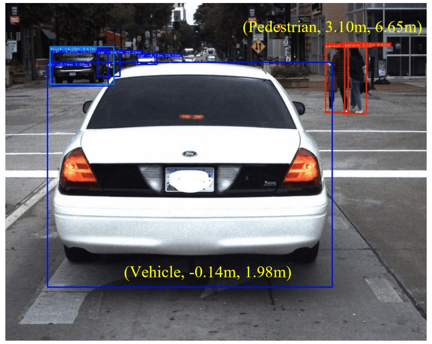
FIGURE 4. A turkey using "evidence"; unaware of Thanksgiving, it is making "rigorous" future projections based on the past. Credit: George Nasr

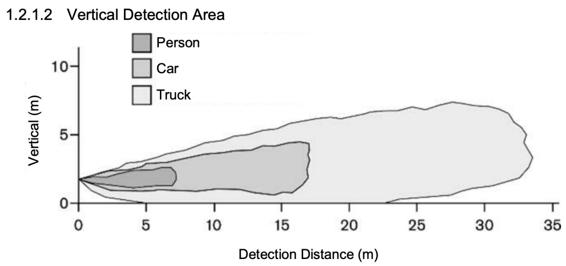


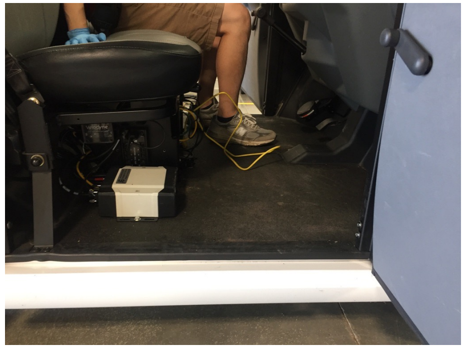
Baked-in Assumptions in our example

- Perception.
  - Sensor detects obstacle **iff** distance $d \leq D_{sense}$
  - No false positives, negatives, probabilities
  - Pedestrian is known to be moving with constant velocity from initial position. This will be used in the safety analysis, but not in the vehicle's automatic braking algorithm
- No sensing-computation-actuation delay.
  - The time step in which $d \leq D_{sense}$ becomes smaller is exactly when the velocity starts to decrease

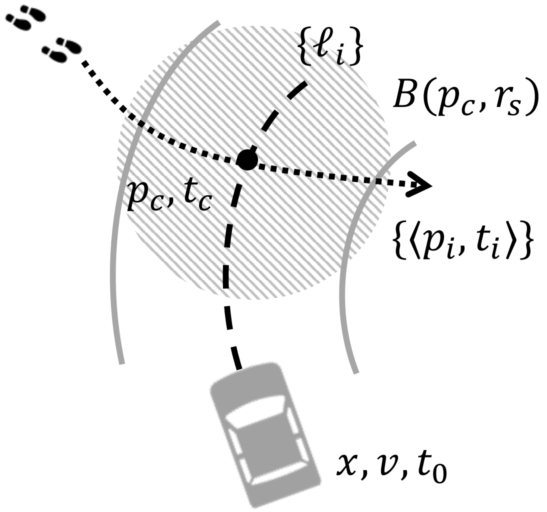








Baked-in Assumptions (continued)



- Mechanical or Dynamical assumptions
  - Vehicle and pedestrian moving in 1-D lane.
  - Does not go backwards.
  - Perfect discrete kinematic model for velocity and acceleration.
- Nature of time
  - Discrete steps. Each execution of the above function models advancement of time by 1 step. If 1 step = 1 second, $x_1(t+1) = x_1(t) + v_1(t).1$
    - We cannot talk about what happens between [t, t+1]
  - Atomic steps. 1 step = complete (atomic) execution of the program.
    - We cannot directly talk about the states visited after partial execution of program

Summary

- Absolute safety checking boils down to showing that none of the executions of the automaton reaches an unsafe set U
- To reason about all executions of we have to work with infinite sets of states
- One way to compute infinite sets is using the Post operator
- But, computing all executions for unbounded time can be hard
- Invariant trick (i) $Q_0 \subseteq I$ and (ii) $Post(I) \subseteq I$ can give a shortcut for proving safety
- The inavariant I may contain important information about conserved quantities, and thus, may tell us why the system is safe, and not just that it is so
- Mind the gap between model and reality
- Next: Perception