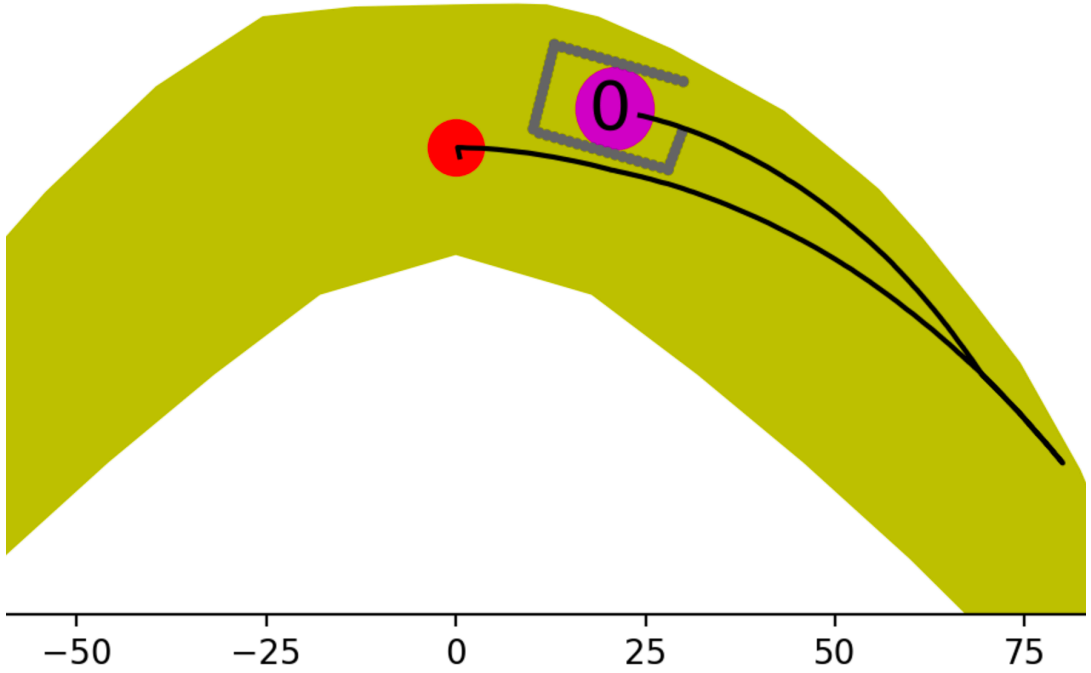
Meeting
05/21/2020

Shuo Zhang

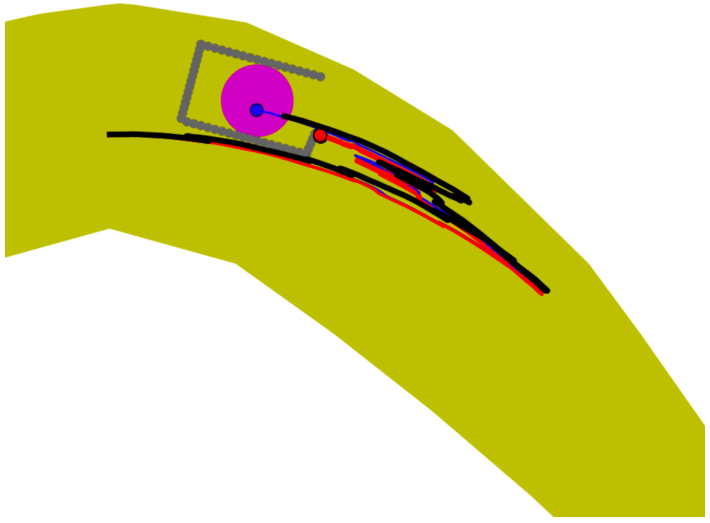
Horsehoe case

dist_sofar+dist_togoal
+10000*check_small_action

Number of same actions: 100
Planning time: 5hours 26mins
Number of expanded nodes: 31729
Maximum depth of tree: 11
Planned path depth: 8



Avishai's planned path also work.
However, his planned path has more depth and needs more time.



PPO RL using Transition model as env to generate data

- Progress:
  1. Wrapped the transition model as a class of sample generation environment
  2. Adjusting my previous PPO code(for gym env) to our case of transition model env
  3. Almost ready to train

Discussions and Questions:
PPO RL using Transition model as env to generate data

- State: (current_state + goal_location + goal_range)? Do we need goal range? Or can we just assume goal_range is some constant, ex. 5mm?

Because some testcases in Avishai's paper used 8mm as goal range, some used 5mm

- When hand reaches within goal range, should we define reward to be 0 or the actual distance to the goal? (ex. Goal range 5mm, when hand is 4.7mm far from the goal, should we define reward to be 0 or -4.7 ?)

- Currently env without any obstacle. Should I implement some fixed obstacles in the env? Or the locations of obstacles are also parts of State? Like goal location?

- I am going to define rewards of failure as ex. "-10000",
  in case of invalid region, invalid gripper load and collision