Meeting 09/07/2020

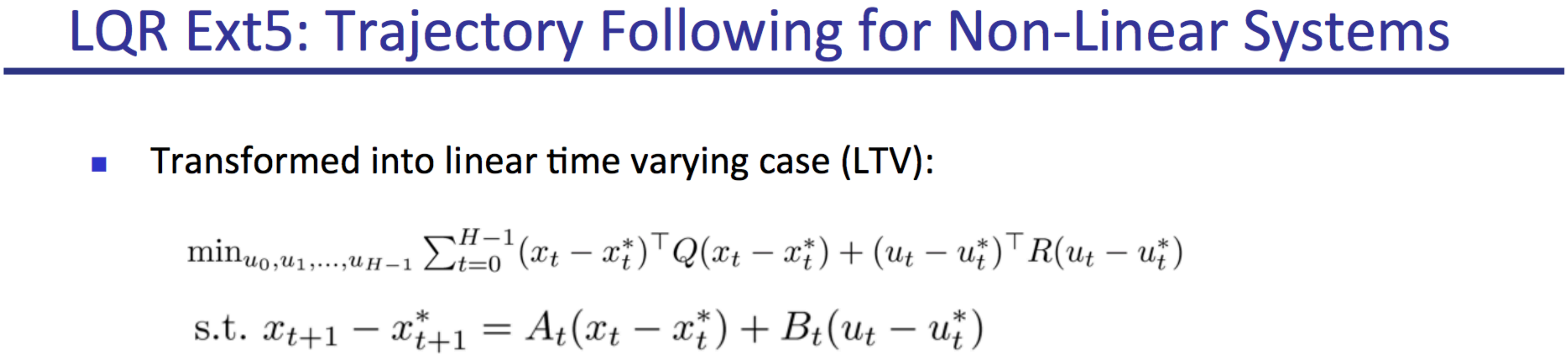
Shuo Zhang

Past week

- Re-implmented LQR with Matrix R=0 rather than Identity Matrix, since we want to follow the trajectory x* rather than the action u*.

- Implemented (A*-based)LQR closed-loop control for Reacher (3 goal locs)

- Calculated Matrix K for all 4 tasks



LQR Ext5: Trajectory Following for Non-Linear Systems

- Transformed into linear time varying case (LTV):

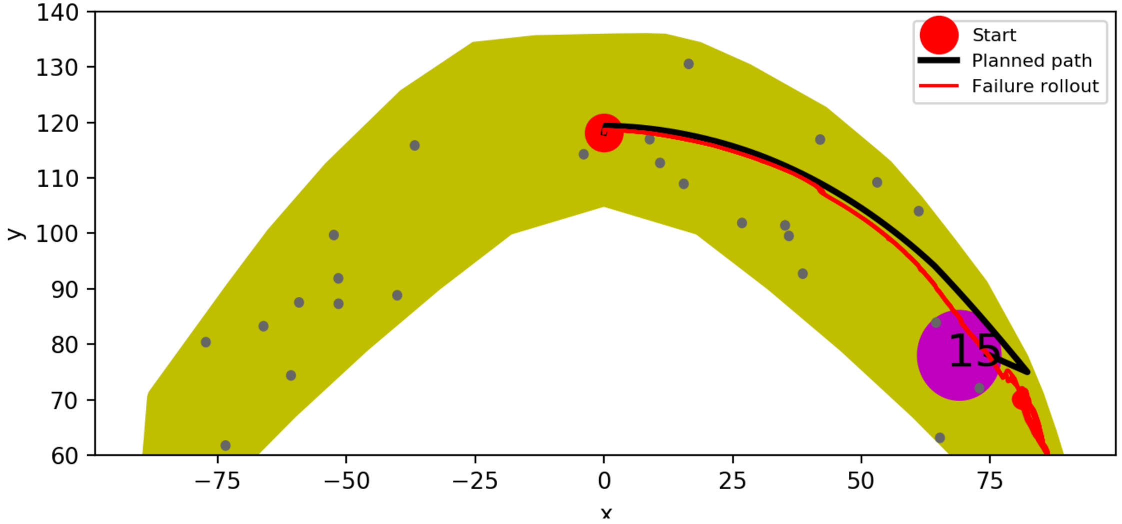
$$\min_{u_0,u_1,\ldots,u_{H-1}} \sum_{t=0}^{H-1} (x_t - x_t^*)^\top Q (x_t - x_t^*) + (u_t - u_t^*)^\top R (u_t - u_t^*)$$

$$\text{s.t. } x_{t+1} - x_{t+1}^* = A_t (x_t - x_t^*) + B_t (u_t - u_t^*)$$

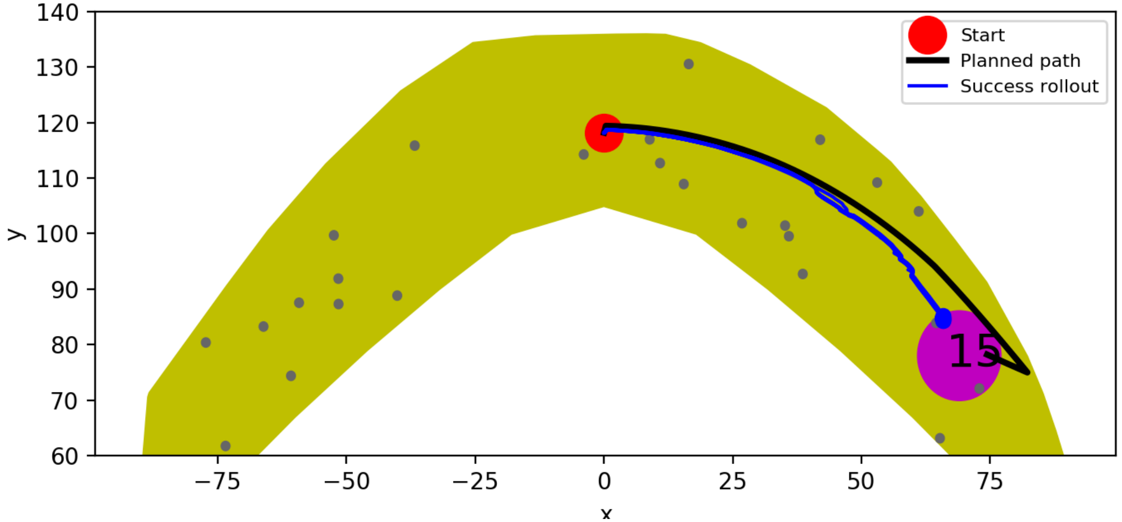
Gazebo Hand: Goal Reach Rate

Goal Location	0	2	7	8	15
A*	0%	100%	100%	0%	0%
LQR(Q=E)	0%	100%	100%	100%	100%
LQR(Q=0)	100%	100%	100%	100%	100%

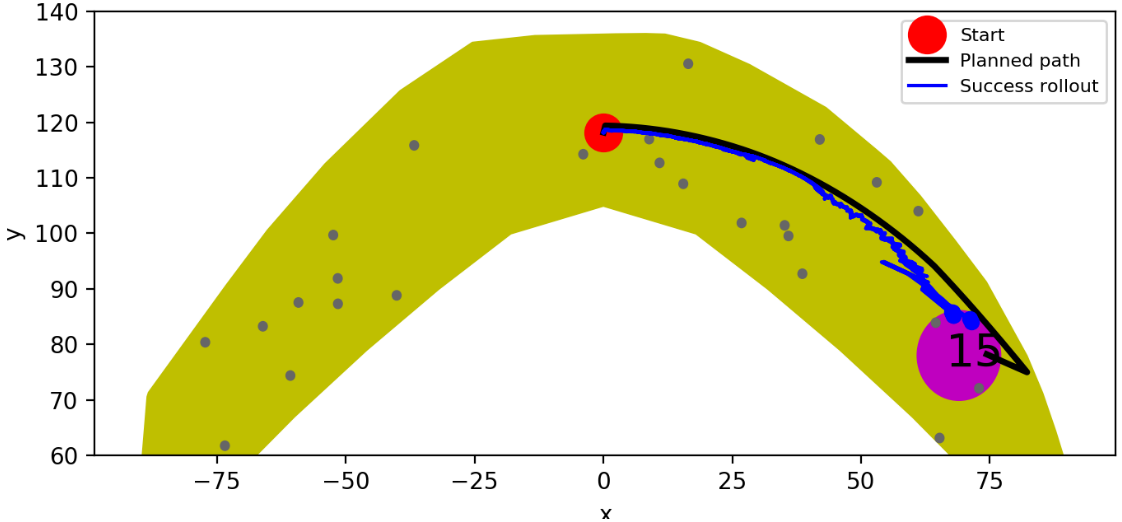
Goal Location 15



A*

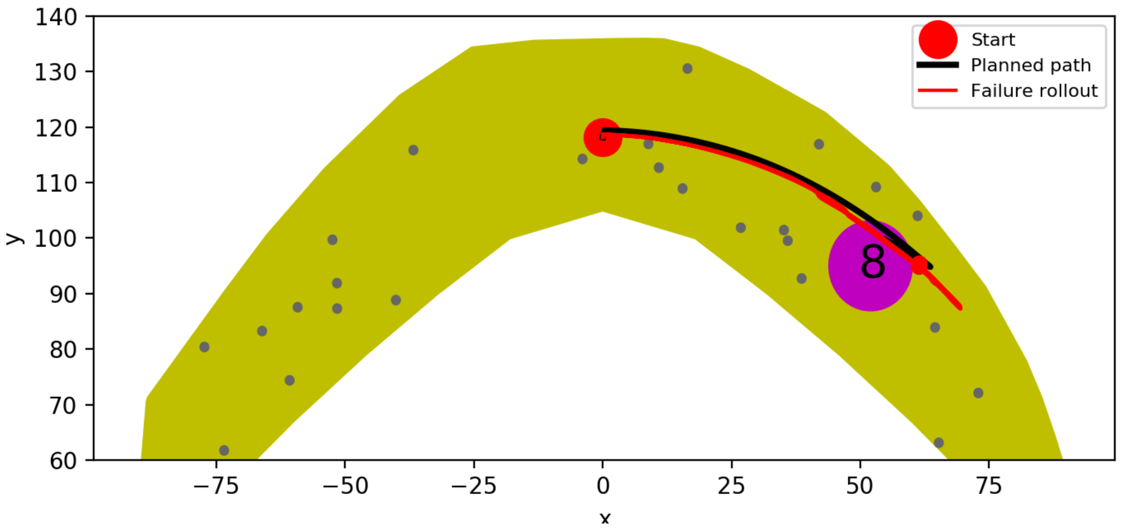


LQR (R=E)

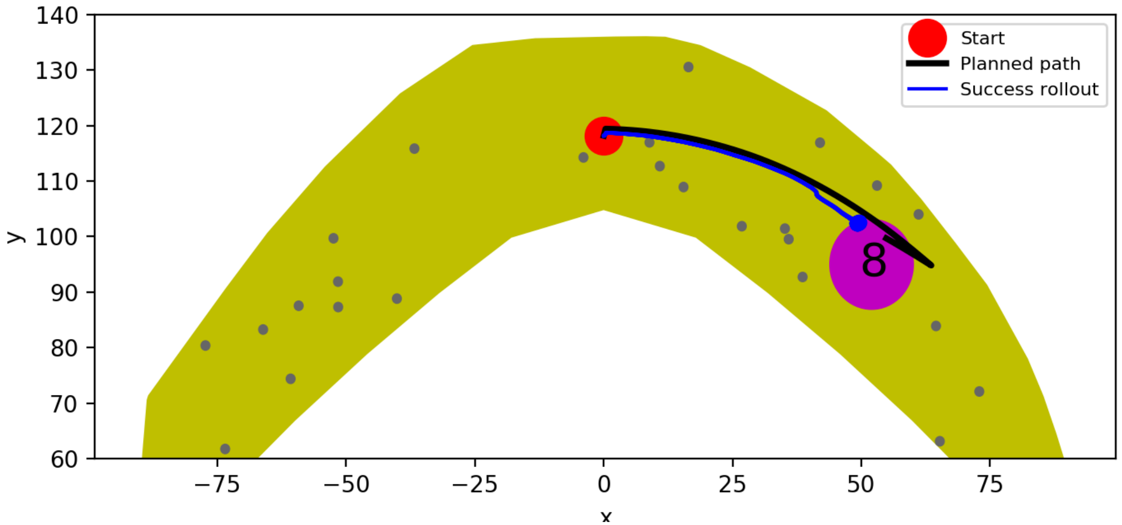


LQR (R=0)

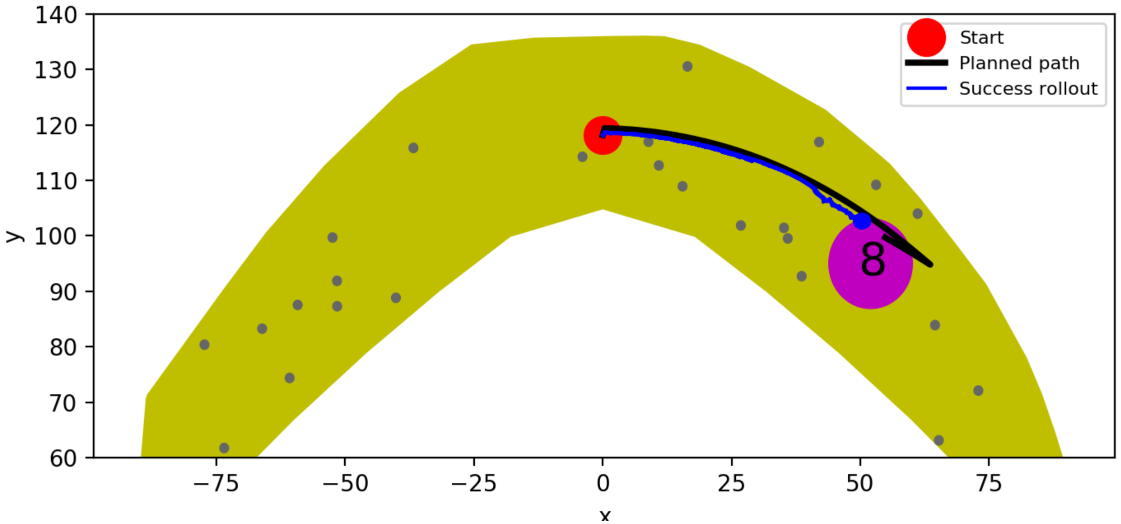
Goal Location 8



A*

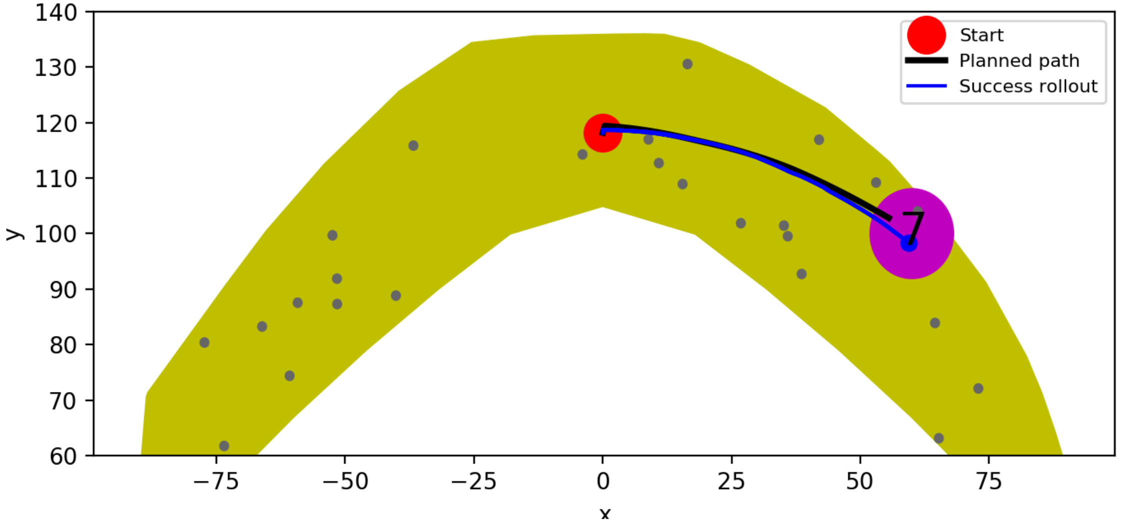


LQR (R=E)

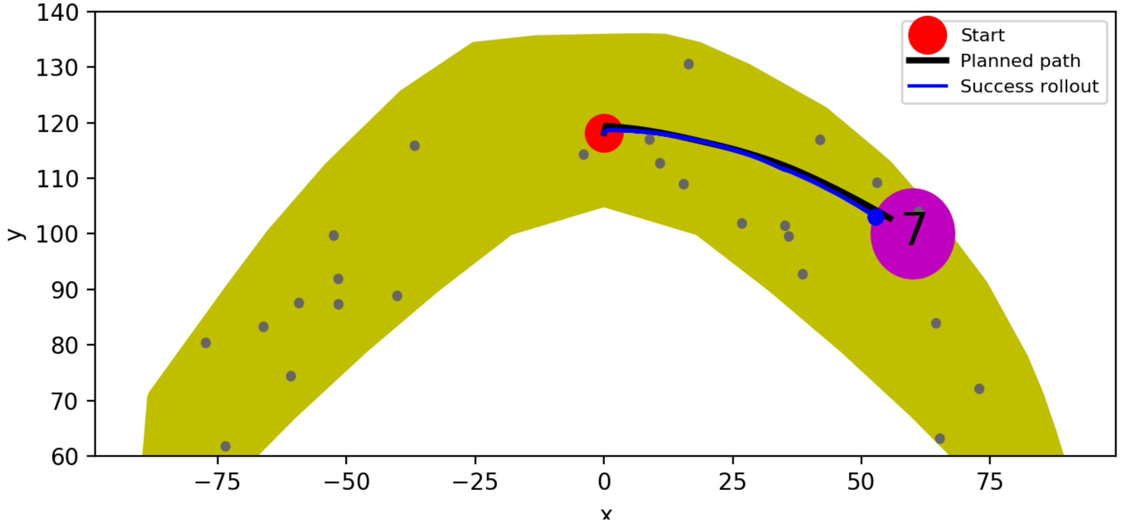


LQR (R=0)

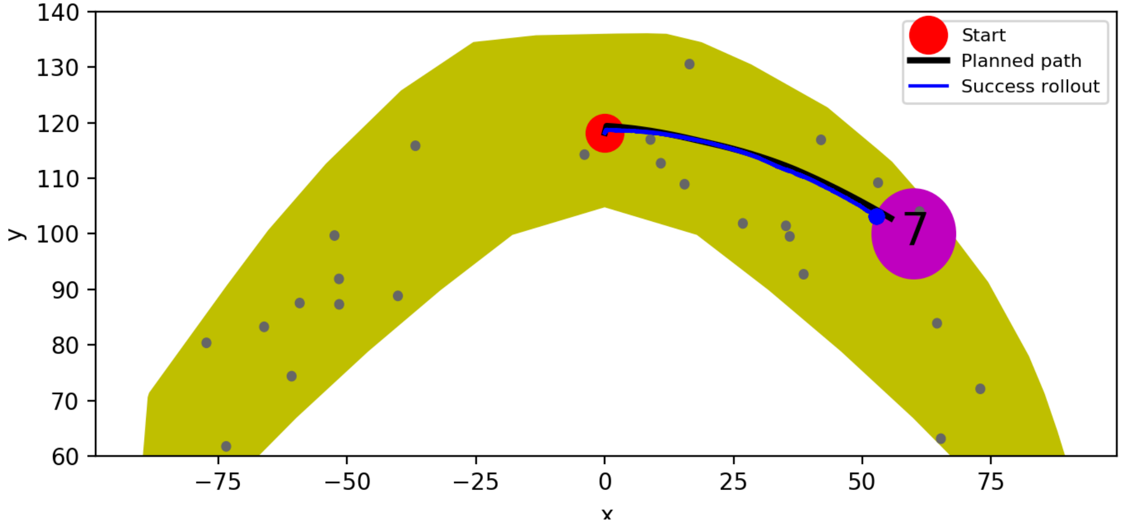
Goal Location 7



A*

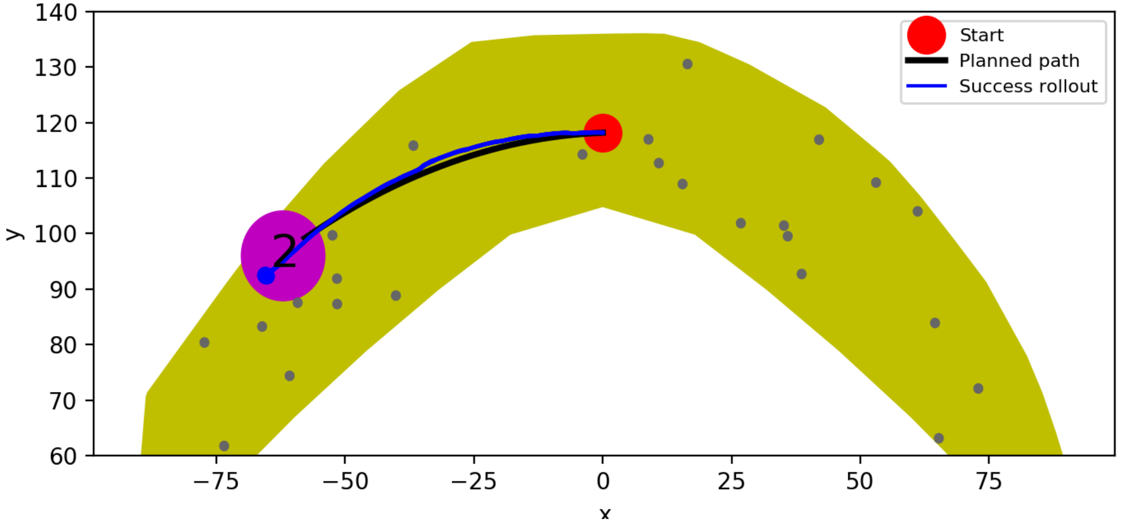


LQR (R=E)

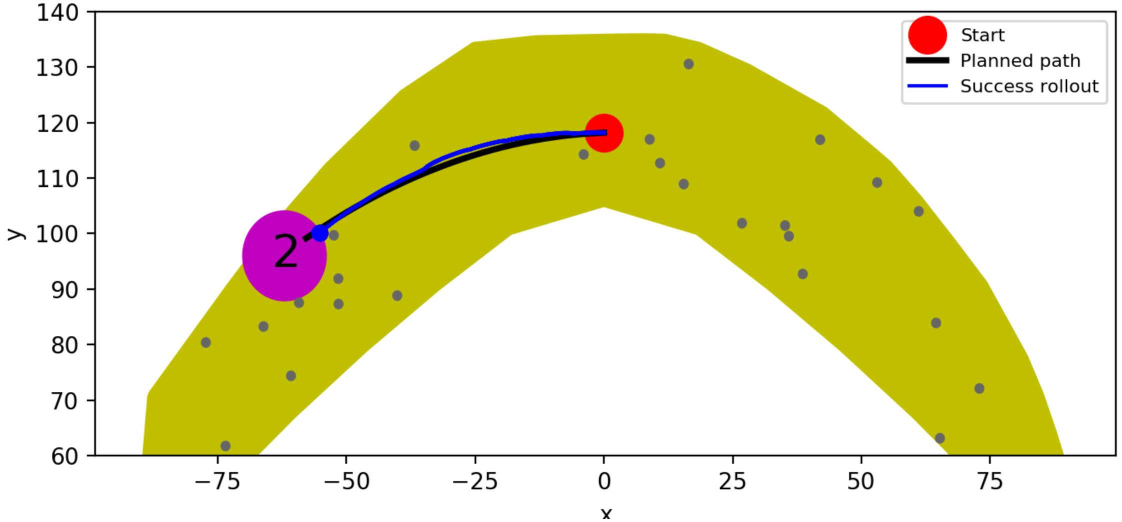


LQR (R=0)

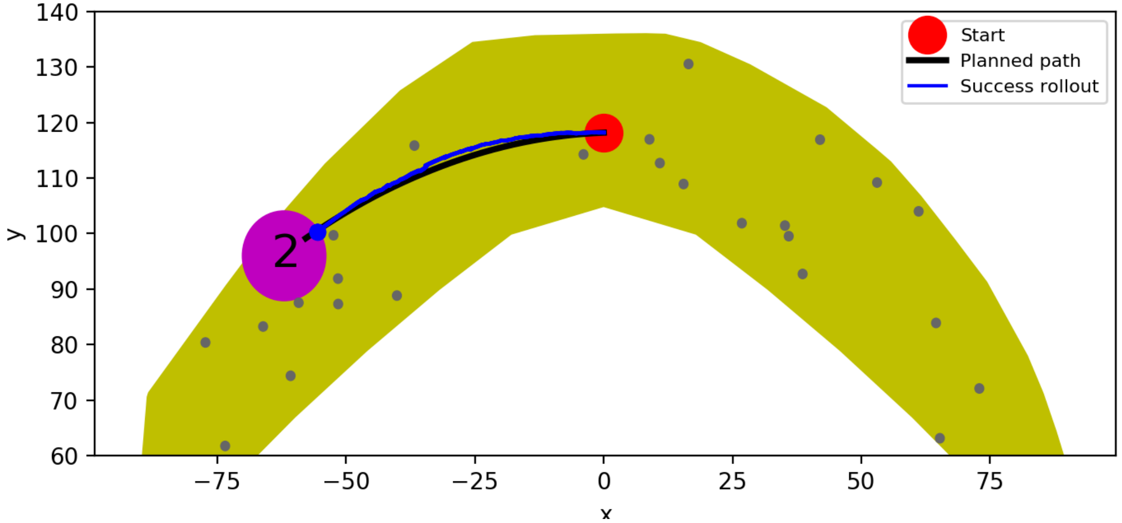
Goal Location 2



A*

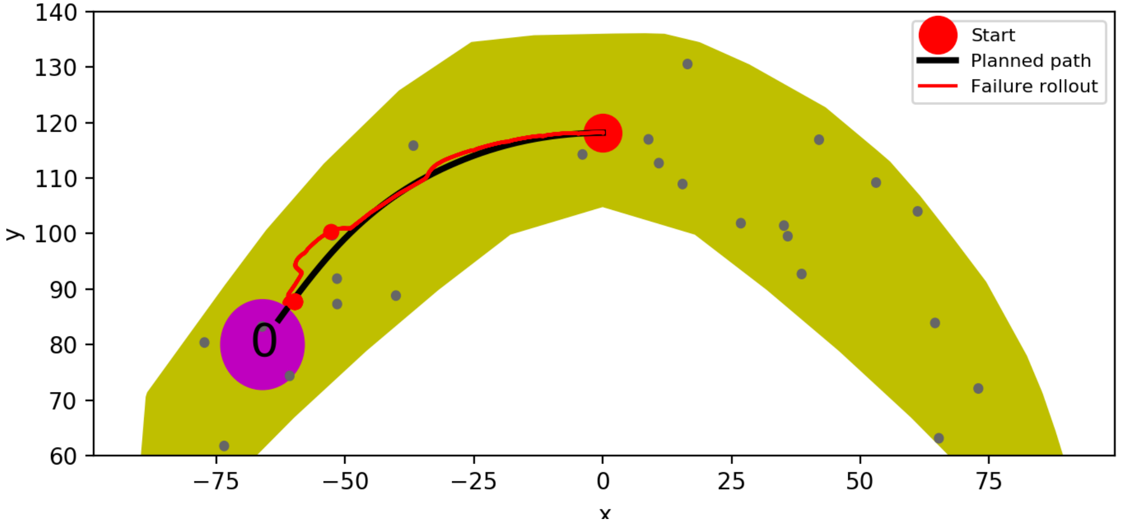


LQR (R=E)

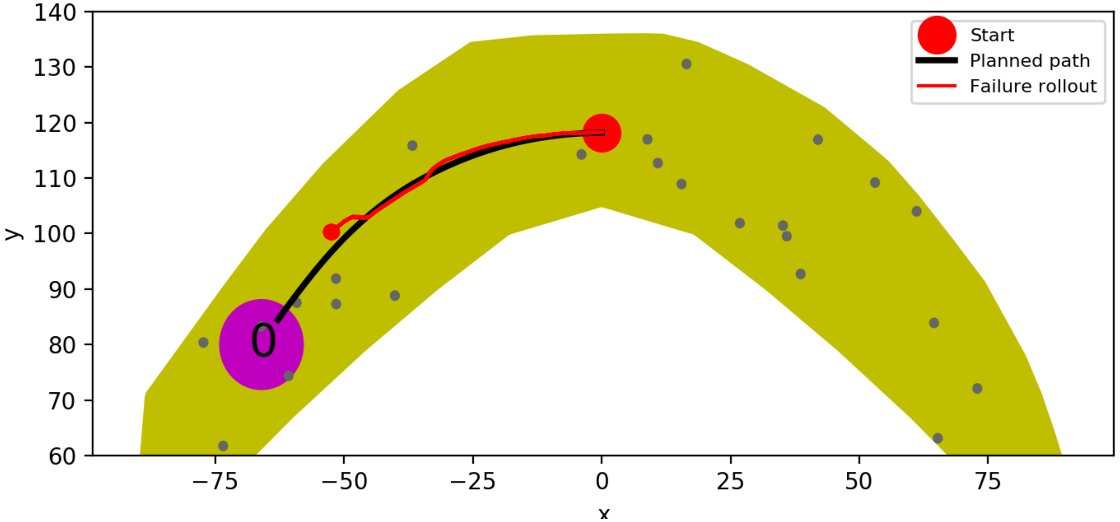


LQR (R=0)

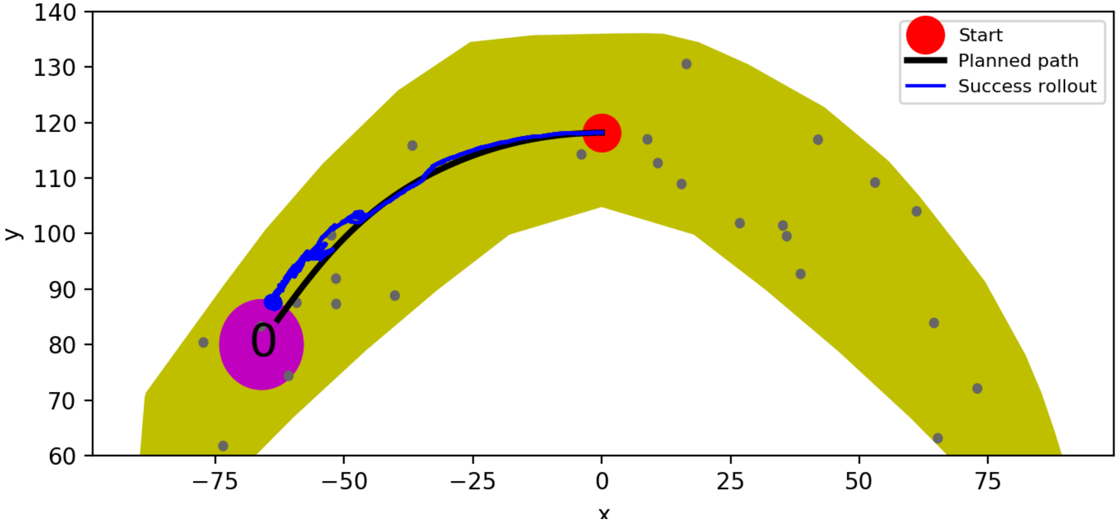
Goal Location 0



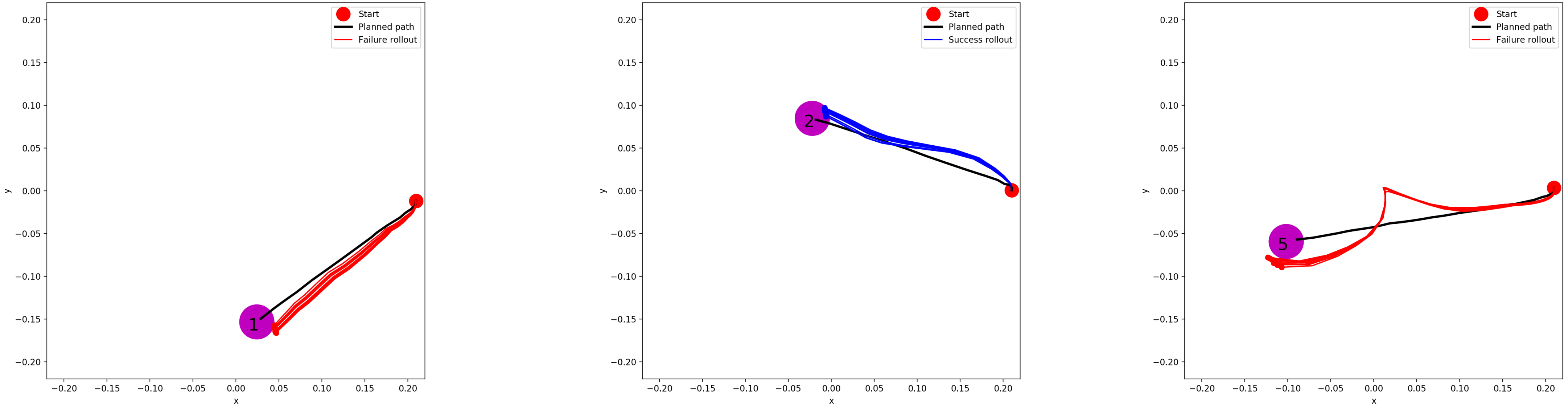
A*



LQR (R=E)

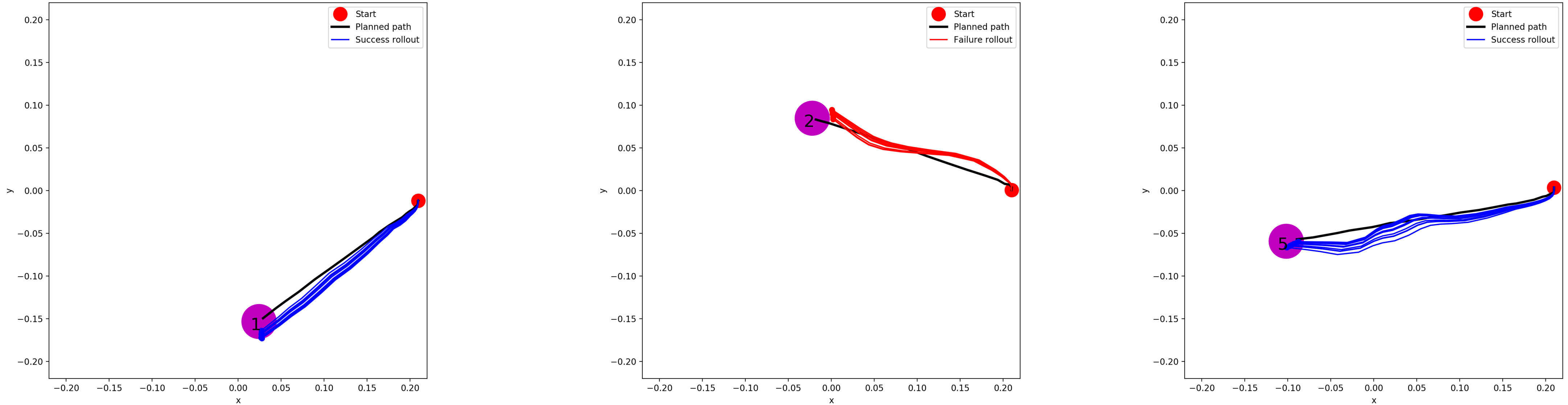


LQR (R=0)



A*

Method	Goal Loc 1	Goal Loc 2	Goal Loc 5
A*	0%	100%	0%
LQR	100%	0%	100%



(A*-based) LQR

Next plans

1) Solve issues of marker tracking. Then, make sure real hand model works with new data. Then, A*+PPO Rollout(100%) on real hand. If too good, reduce data+retrain dynamics+redo A* and PPO and their rollouts? Then, LQR.

2) Make gazebo hand task more difficult? (e.g. 1% number of trajectories, rather than data size)

3) LQR on Acrobot? Switch to other Mujoco tasks?

4) Derive new equations, objective function and optimization theory for AIP

5) Implement AIP

6) Should also try closed loop control using PPO (trained from model) on real environment?

AIP Implementation (against TRPO,PPO)

Difference 1 Policy Network:

TRPO/PPO: u_final=pi_theta([x])

AIP: u_final=pi_theta([x, u_controller])

Difference 2 Constraint:

TRPO:

KL(pi_theta_new || pi_theta_old)<epsilon. (CG+Line Search)

PPO:

No constraint

Constraint (KL(pi_theta_new || pi_theta_old)<epsilon) combined into objective function

AIP:

KL(pi_theta_new || pi_theta_old)<epsilon

??KL(pi_theta || controller)<omega??

(Need to derive new equations and objective for optimization?)

Task	Open Loop A* +Rollout	Open Loop PPO + Rollout	Closed Loop PPO	Cloes Loop LQR based on A*	AIP (3 options)
Reacher (0.1% model)	Done + Done	Done + Done	Not yet	Done	Not yet
Gazebo Hand (0.1% model)	Done + Done	Done + Done	Not yet	Done	Not yet
Acrobot (100% model)	Done + Done	Done + Done	Not yet	Not yet	Not yet
Real Hand (100% model)	Done + Not yet	Done + Not yet	Not yet	Not yet	Not yet