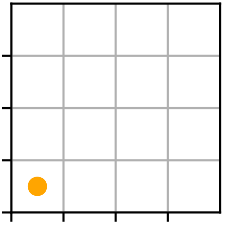
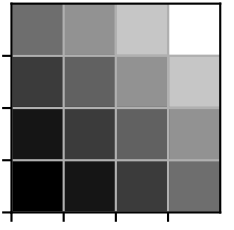
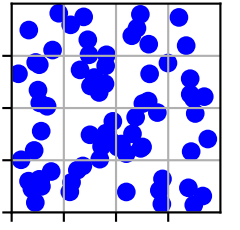
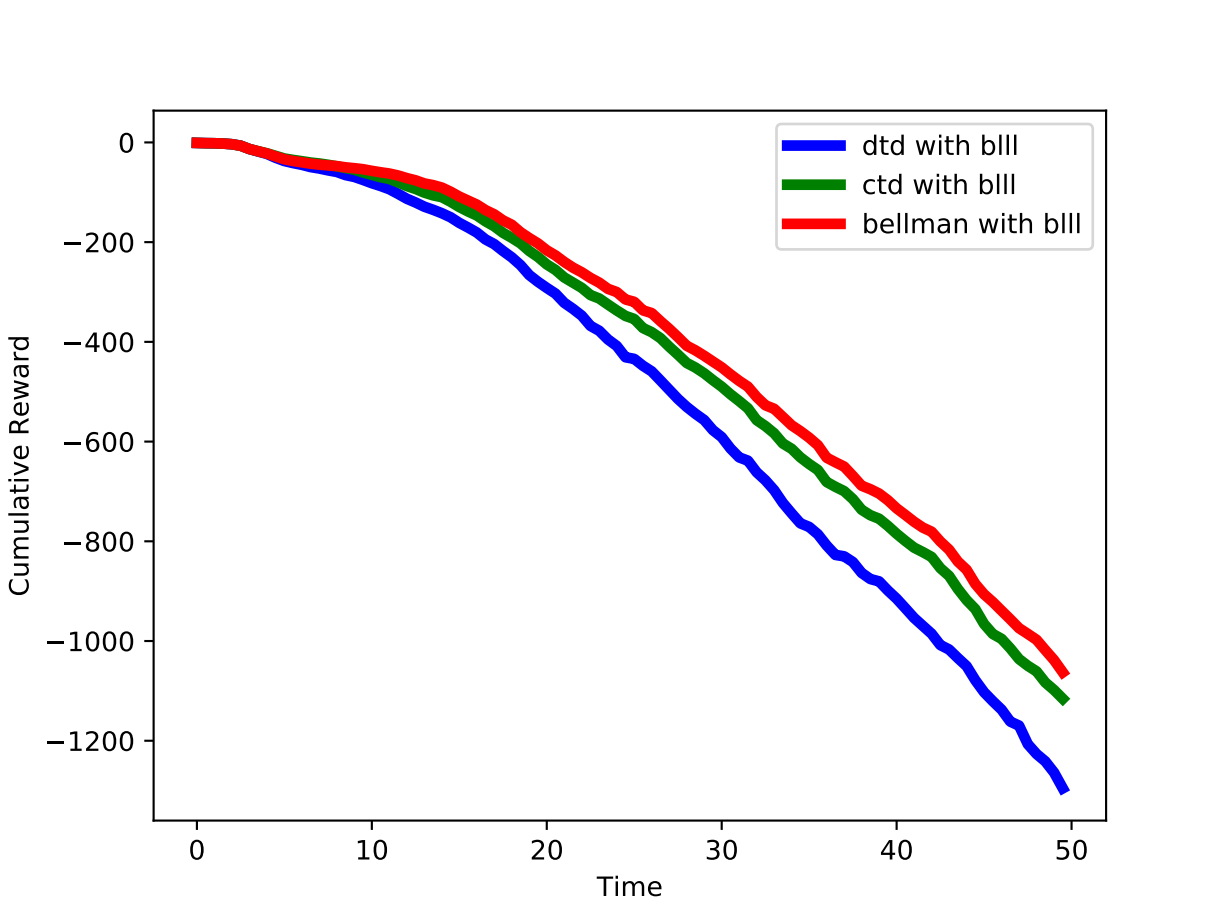
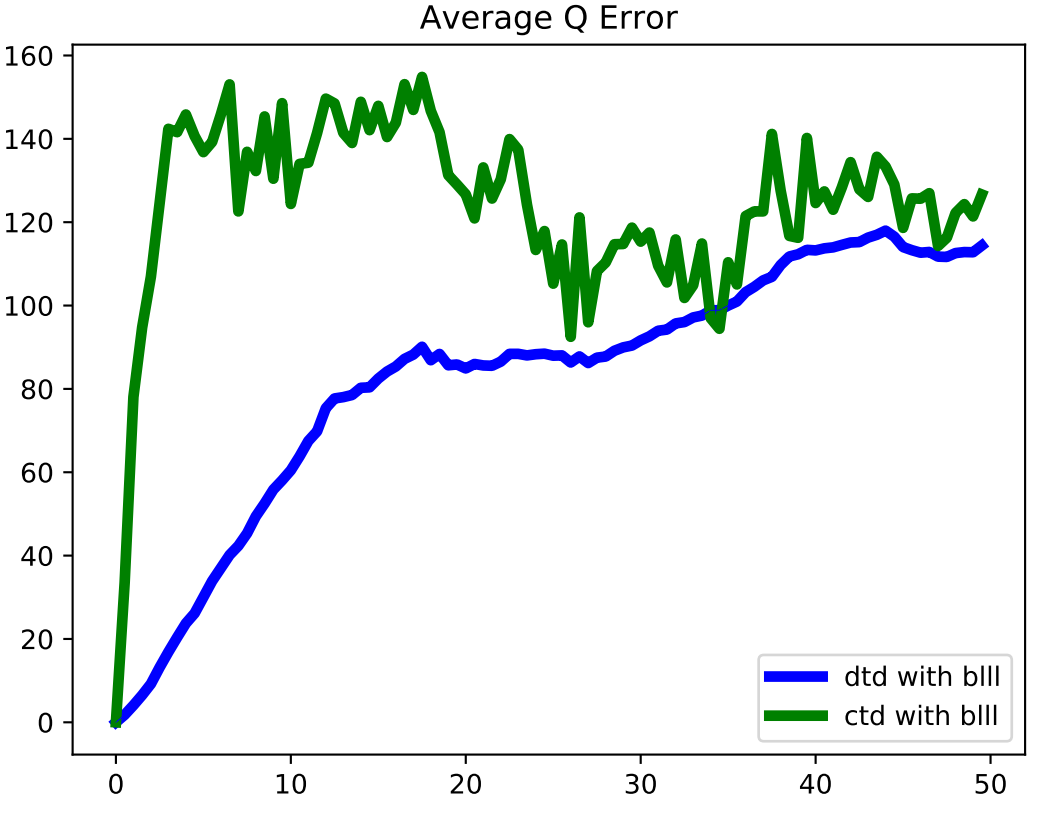
dtd with blll at t/T = 0.0/50.0








dtd with blll
ctd with blll
bellman with blll
Cumulative Reward
Time
0
−200
−400
−600
−800
−1000
−1200
0
10
20
30
40
50


Average Q Error
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
dtd with blll
ctd with blll