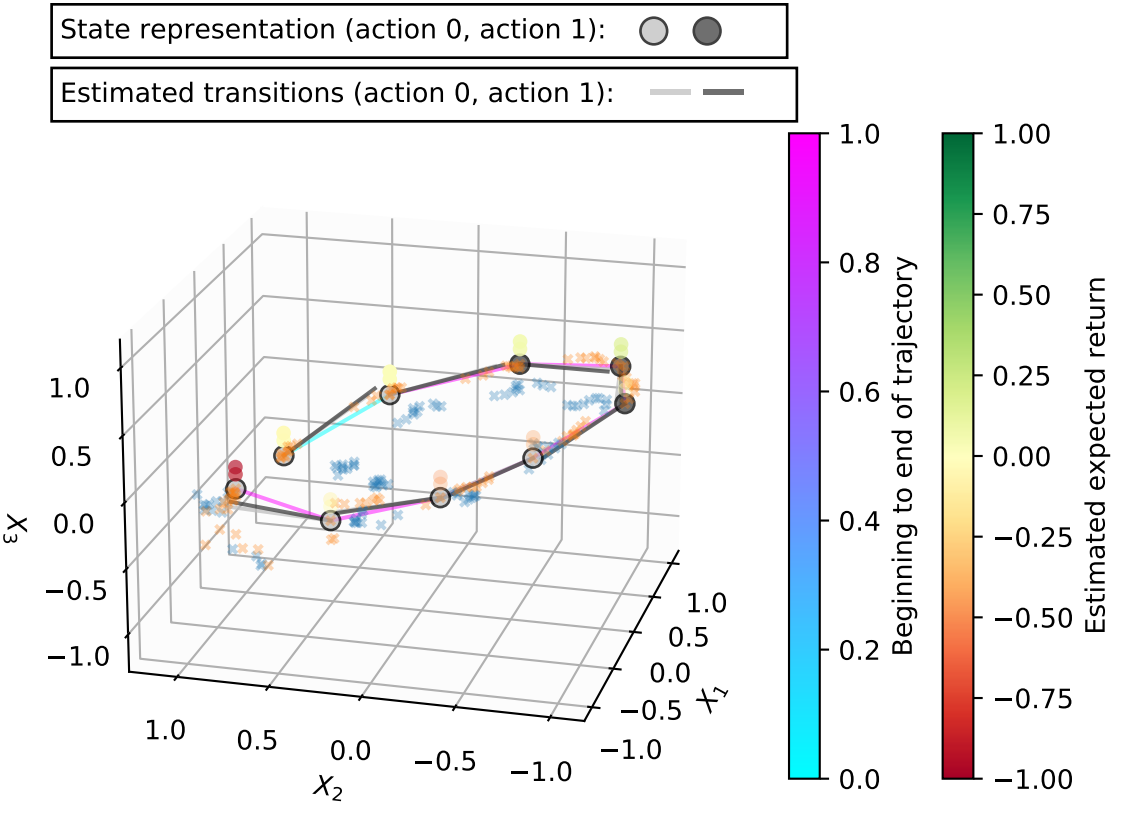
State representation (action 0, action 1):
Estimated transitions (action 0, action 1):
1.0
0.5
0.0
−0.5
−1.0
X_3
X_2
X_1
Beginning to end of trajectory
1.0
0.8
0.6
0.4
0.2
0.0
Estimated expected return
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00