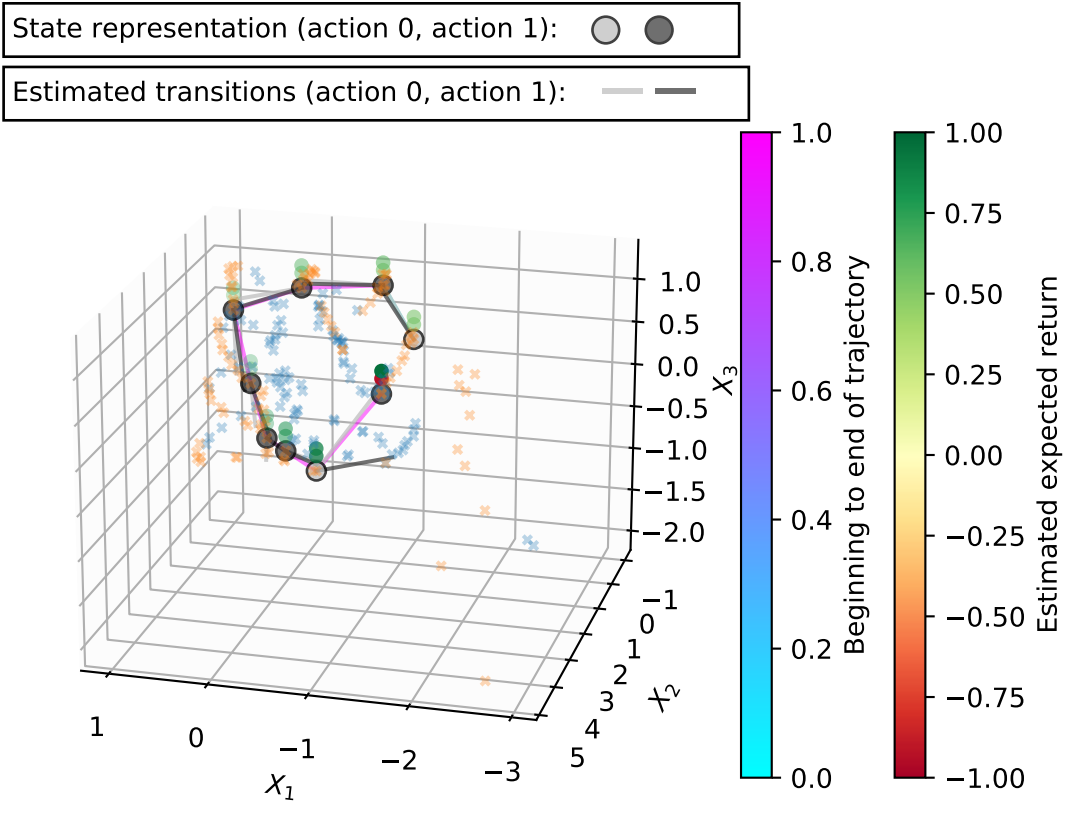
State representation (action 0, action 1):
Estimated transitions (action 0, action 1):
x_1
x_2
x_3
Beginning to end of trajectory
Estimated expected return