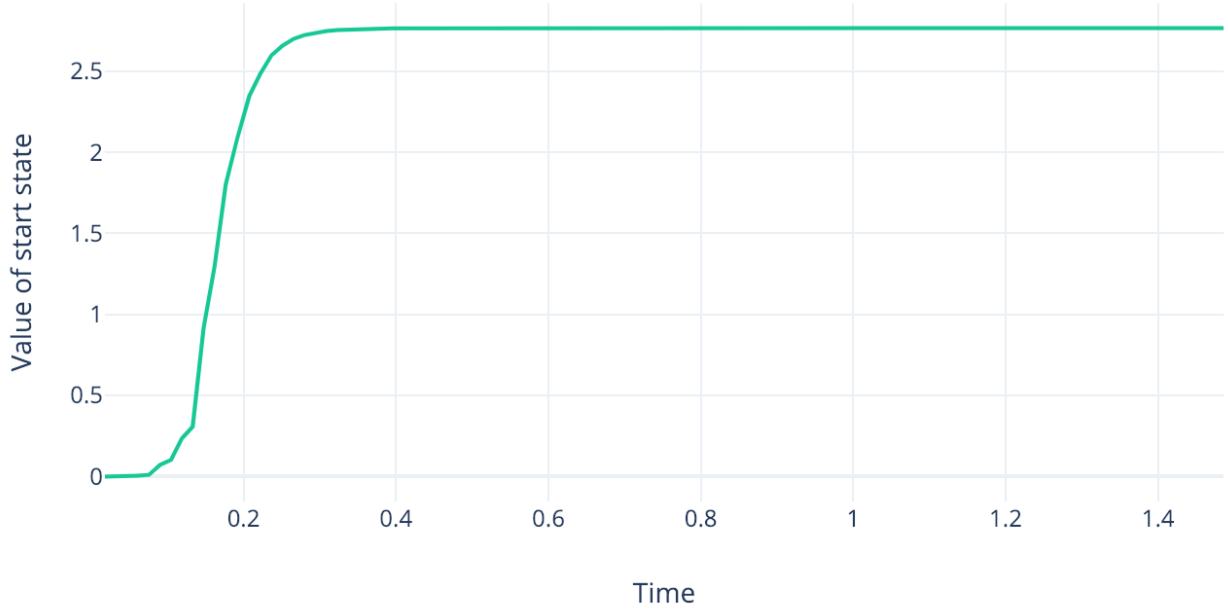
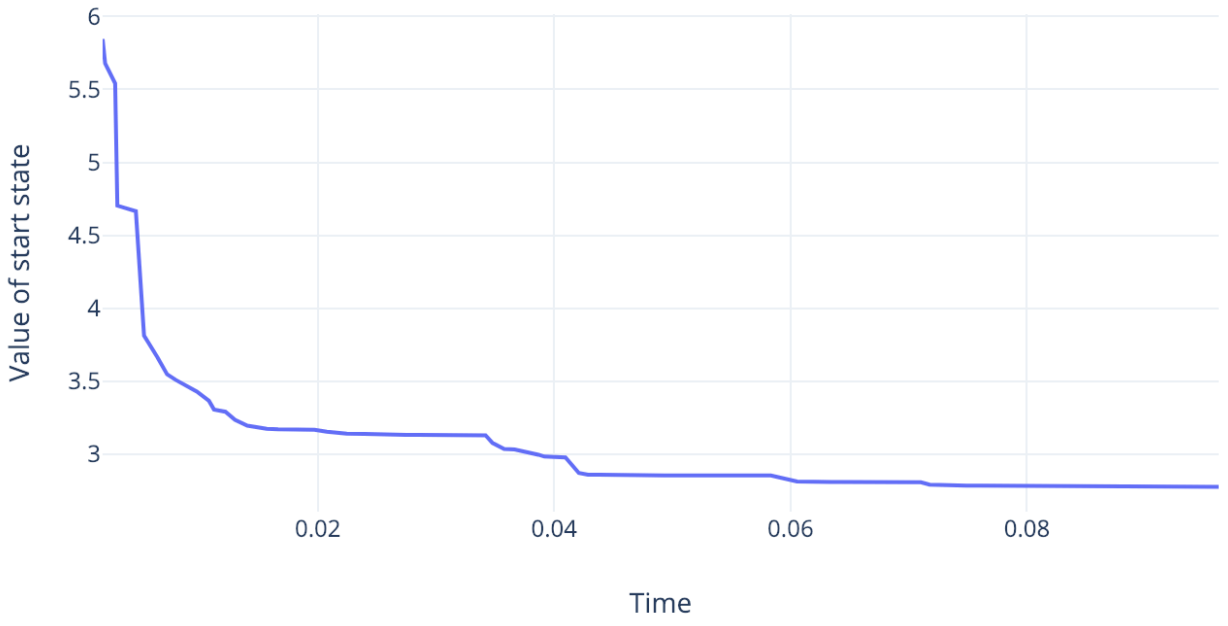
RTDP

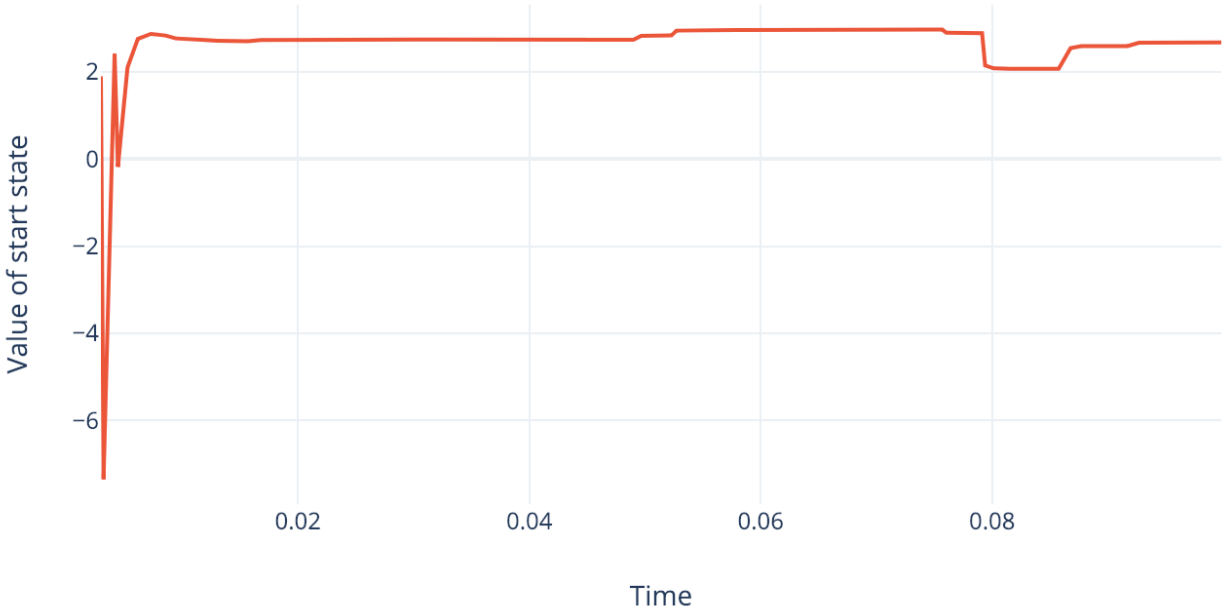
1) Value iteration over time



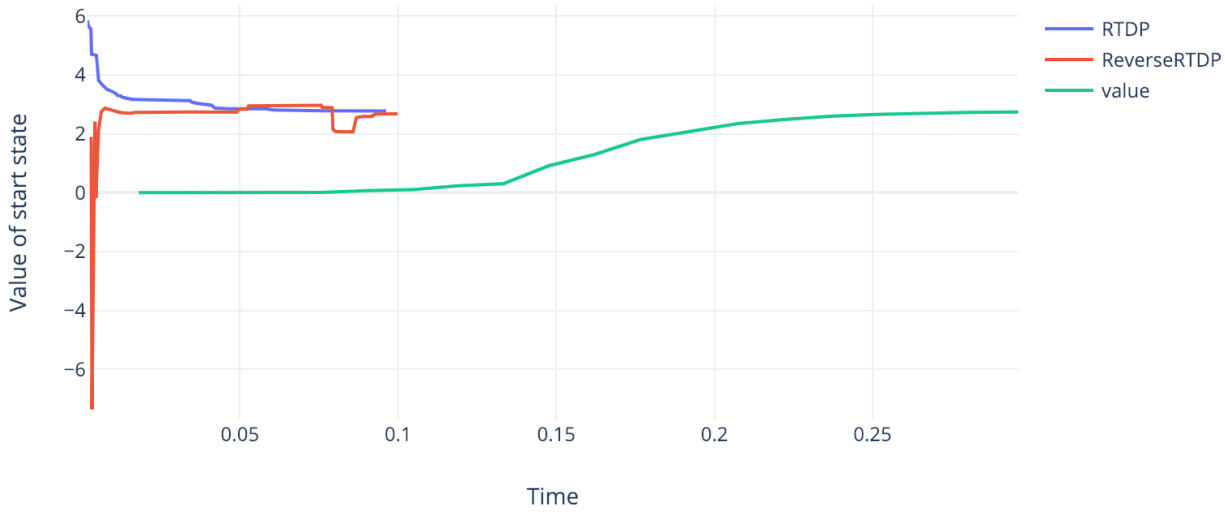
2) RTDP over time



3) Reverse RTDP over time



4) All 3 over time



Analysis:

Reverse RTDP seems to converge the fastest. This is because doing RTDP in reverse places the value updates in the order of effect. Therefore skipping the need for future iterations to correct for the previous iteration update. RTDP is faster than value iteration

My heuristic was discount^(manhattan distance)*reward. This is basically the way reward calculation happens. If we are 3 away then the reward we get is discount^3 of the reward. this is admissible because it doesn't account for other rewards in the system.