Task 6.1

Development of the state is visualised in Figure 1.

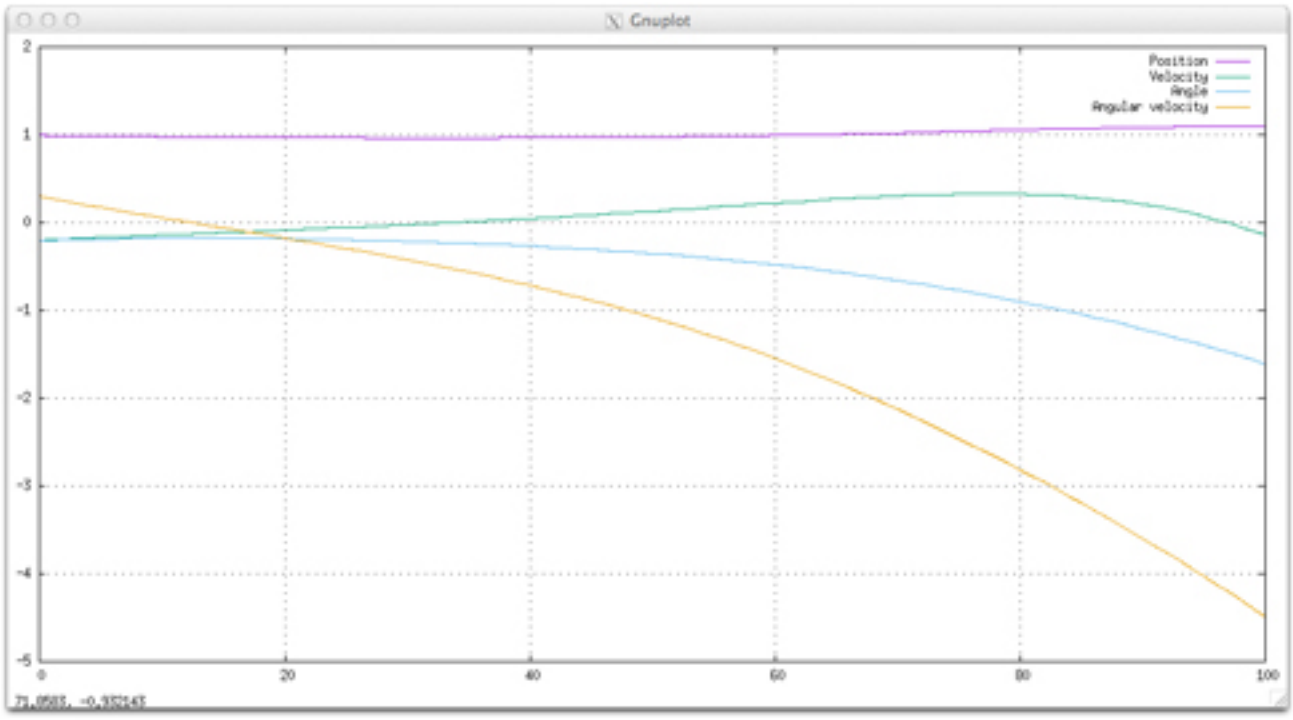

Figure 1: State development with F=0

Task 6.2

Development of the state for parameters $k_1 = -1, k_2 = 3, k_3 = -1, k_4 = 2$ is visualised in Figure 2.

Task 6.3

The learning process was held for 5000 iterations with $\epsilon = 0.8$ and $\alpha = 0.01$. The learning process is visualised in Figure 3.

It can be seen, that from some moment learning process stops improvement of the reward, but from next oscillations of the reward one can be chosen as the best: -994. That reward is got for parameters $k_1 = -485.44374999999945, k_2 = -513.5300000000018, k_3 = -230.34125000000085, k_4 = 281.4124999999998$ and the length of the episode is 646 steps and development of states for these values is visualised in Figure 4.

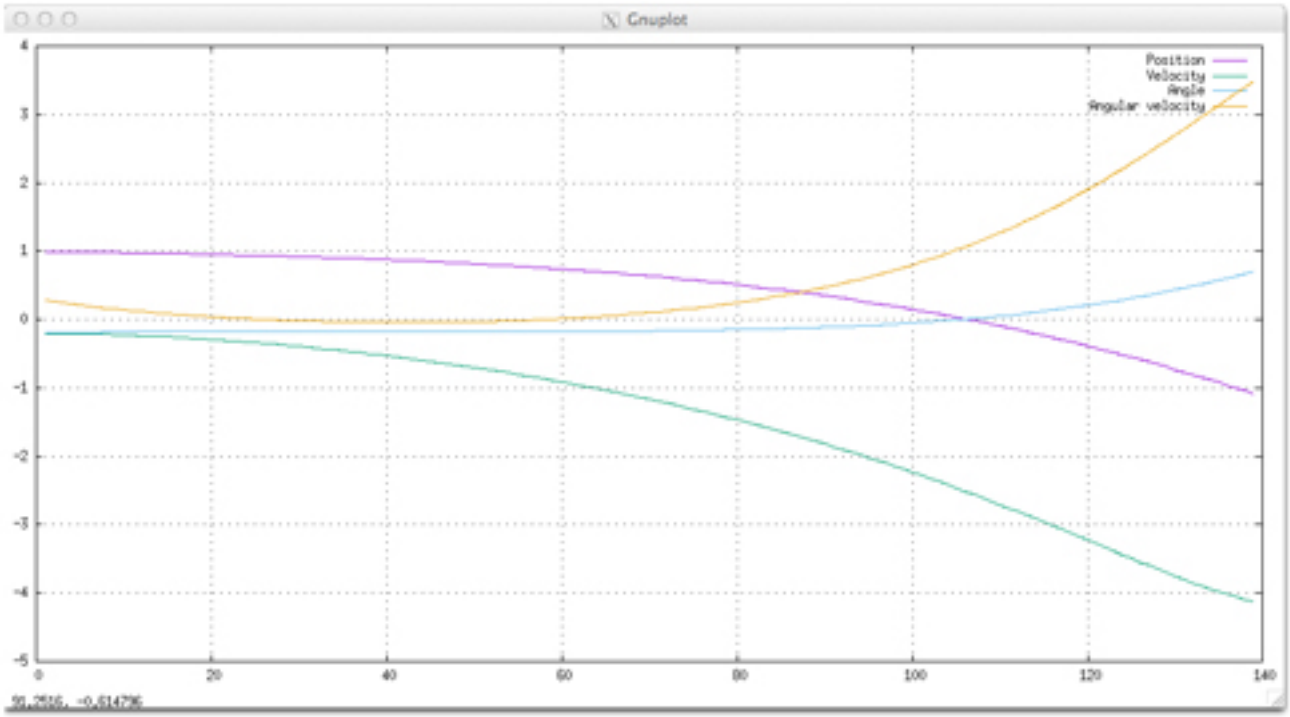

Figure 2: State development with $F = min(20, max(-20, k_1 * position + k_2 * velocity + k_3 * angle + k_4 * angular - velocity))$

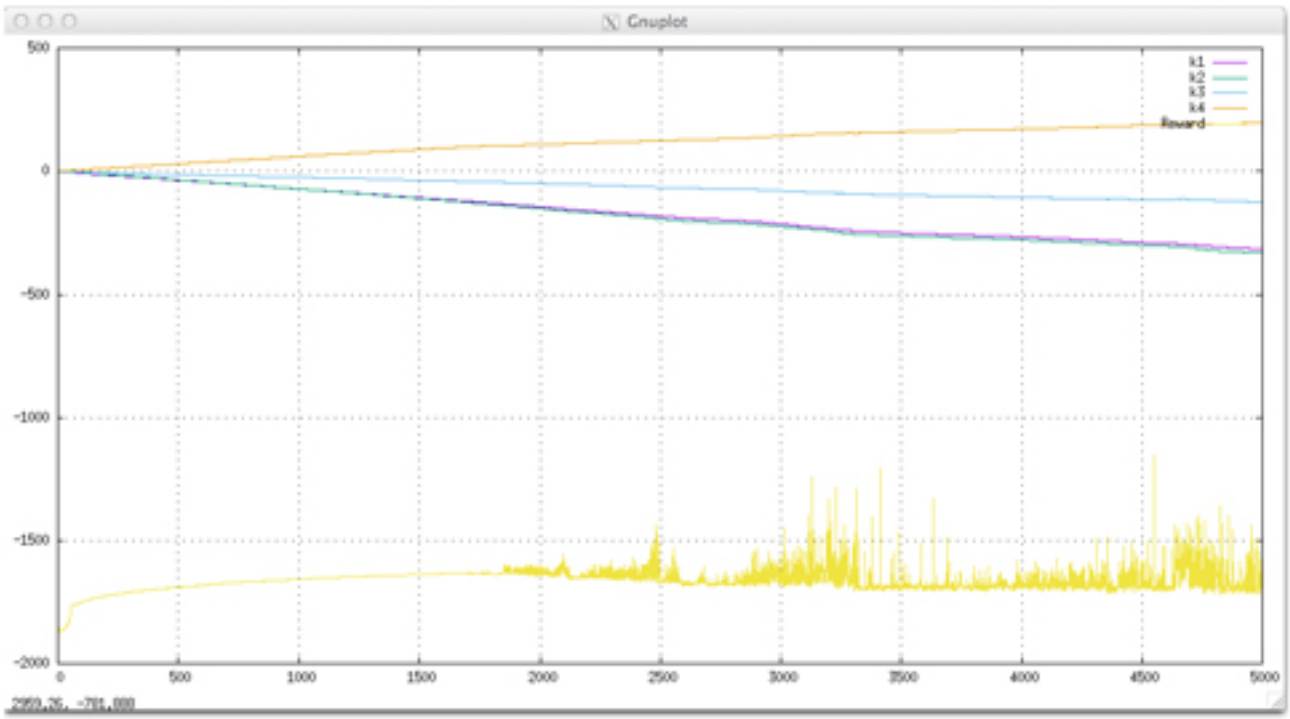

Figure 3: Learning process

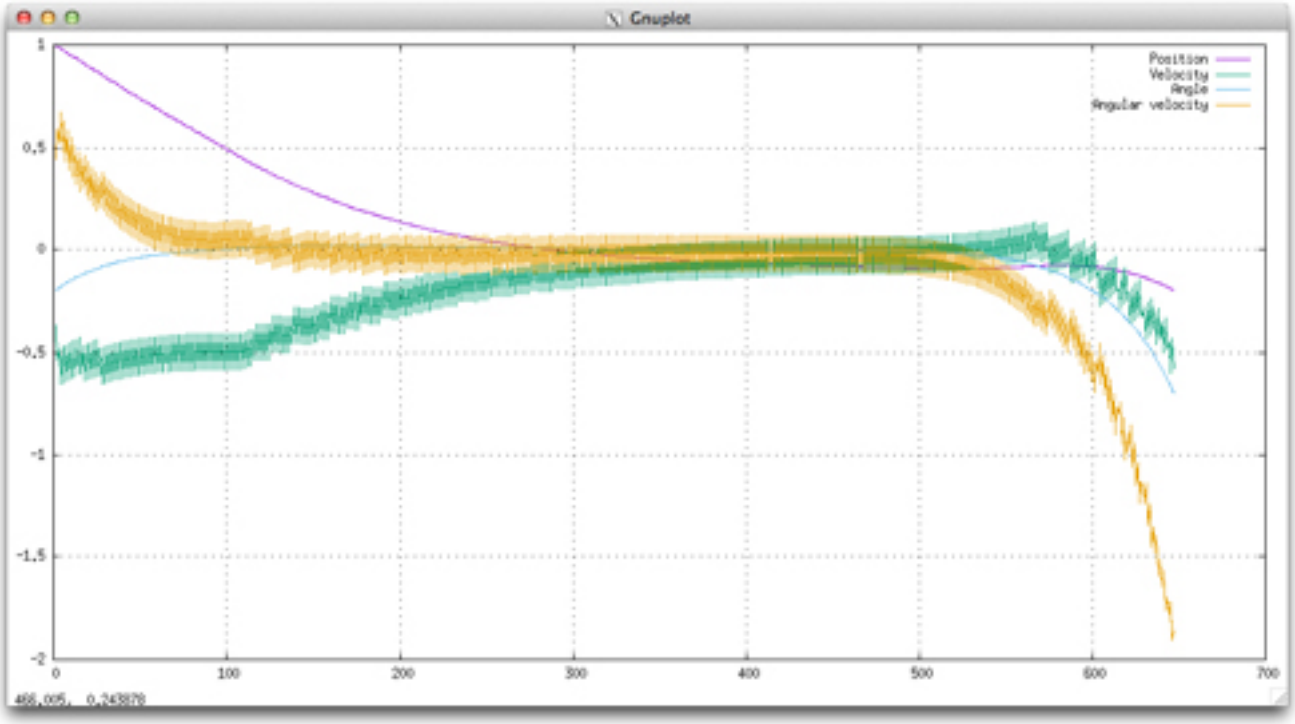

Figure 4: State development for best parameters