RL exercise 08

Mengnan Wang

June 20, 2017

1 Policy gradient method

1.1 Vanilla policy gradient

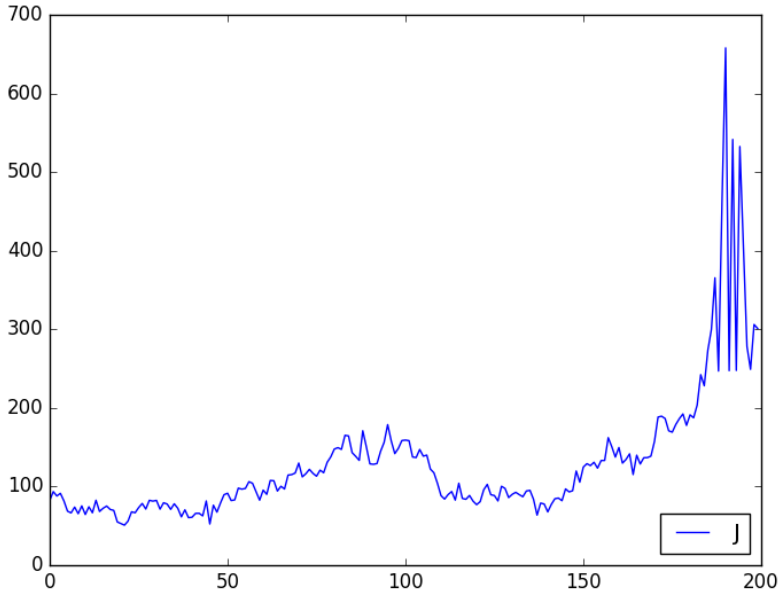

Figure 1: Vanilla policy gradient with fixed step size $alpha = 0.5$

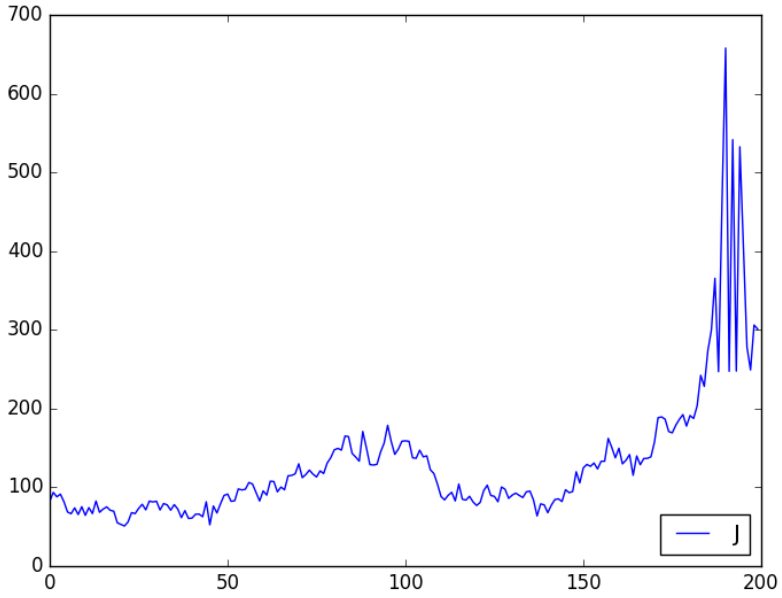

Figure 2: Vanilla policy gradient with existential decay step size $alpha = 10 \exp_{-t}$

1.2 REINFORCE

1.3 G(PO)MDP

2 Learning from demonstrations

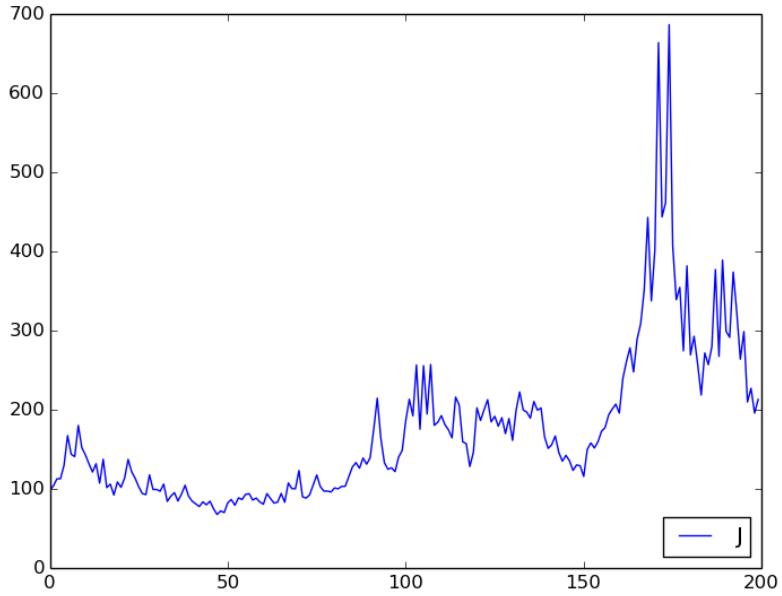

Figure 3: Vanilla policy gradient with Rprop $alpha = 0.5$

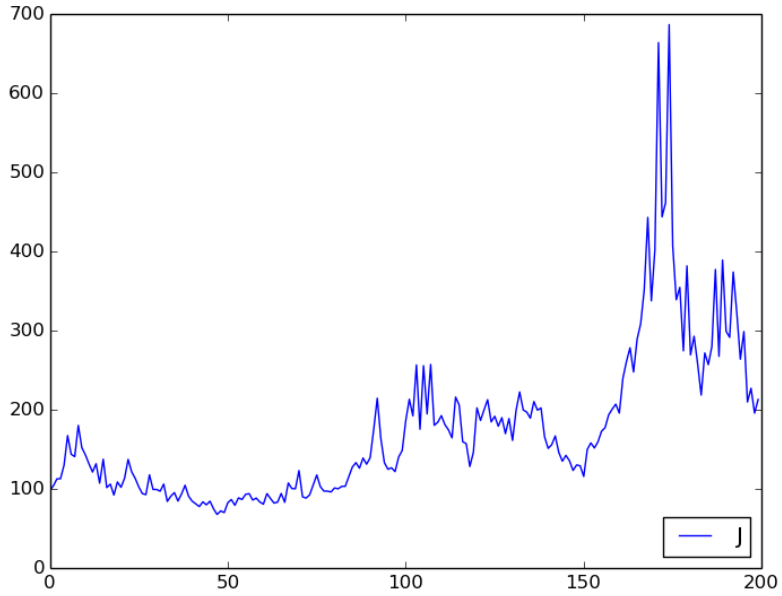

Figure 4: REINFORCE with Rprop $alpha = 0.5$

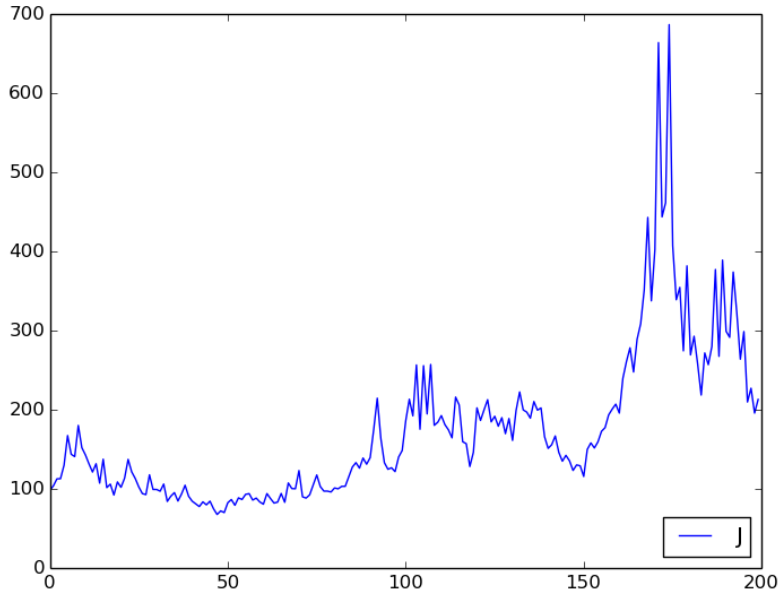

Figure 5: REINFORCE with Rprop $alpha = 0.5$

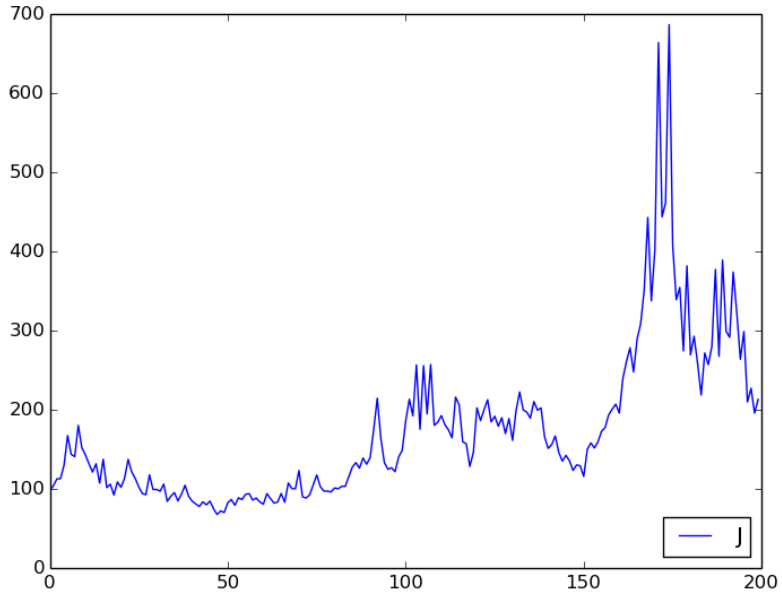

Figure 6: REINFORCE with Rprop $alpha = 0.5$

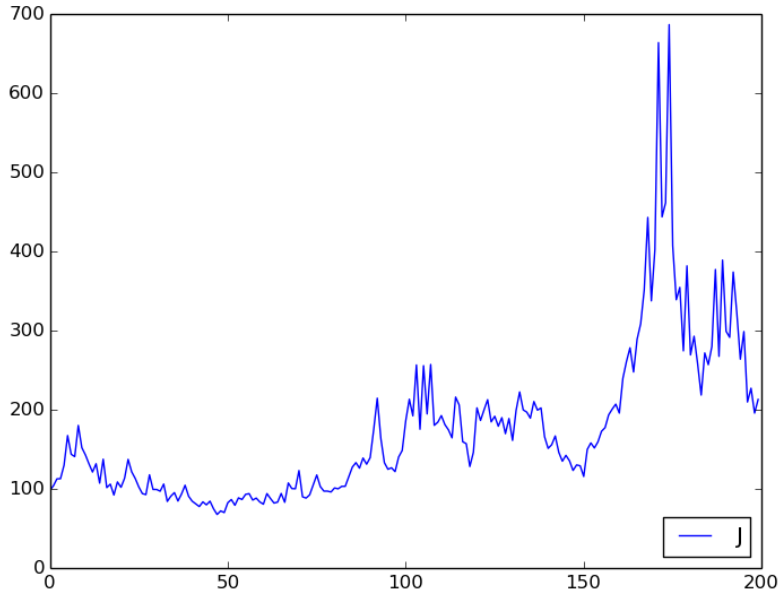

Figure 7: REINFORCE with Rprop $alpha = 0.5$

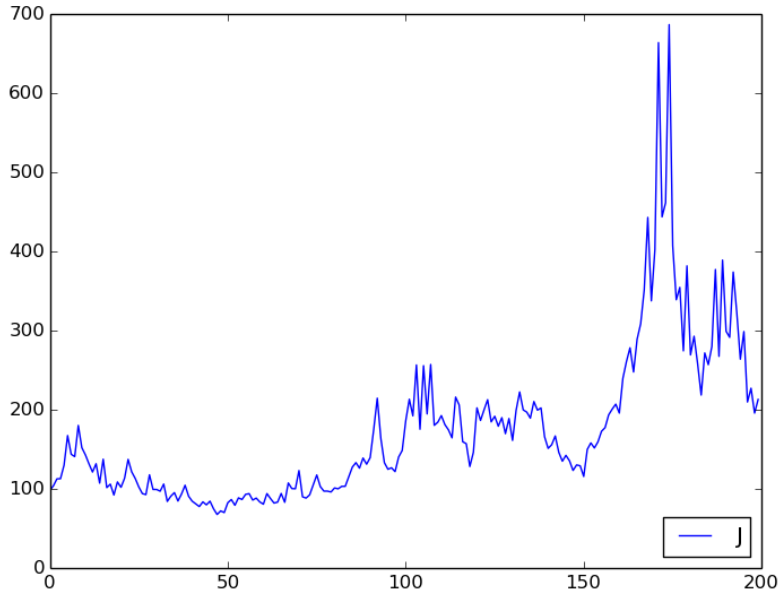

Figure 8: REINFORCE with Rprop $alpha = 0.5$

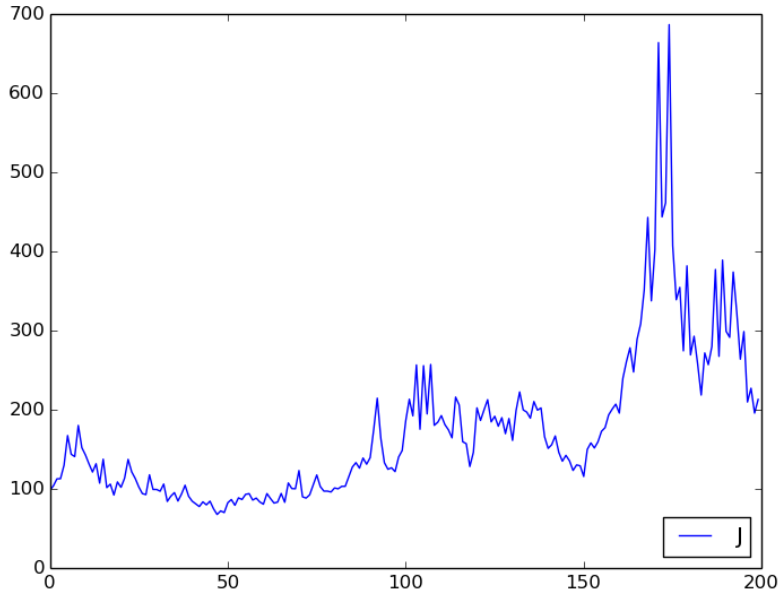

Figure 9: REINFORCE with Rprop $alpha = 0.5$

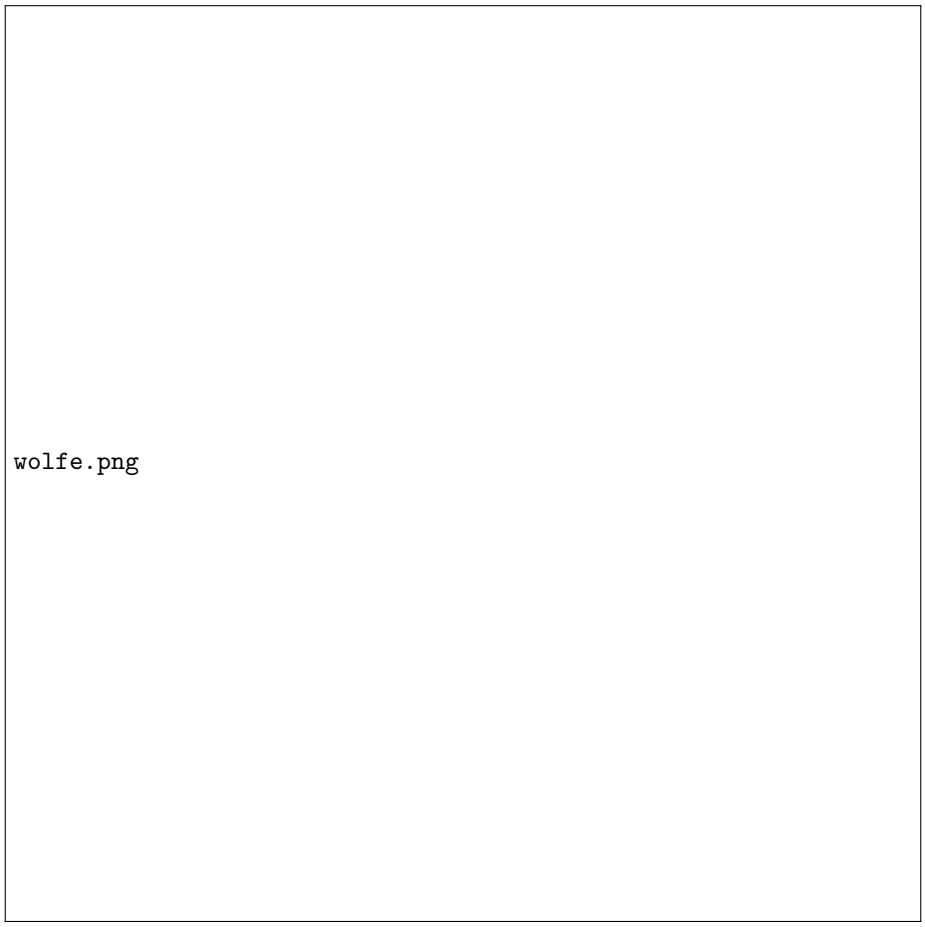

Figure 10: Wolfe condition and line search

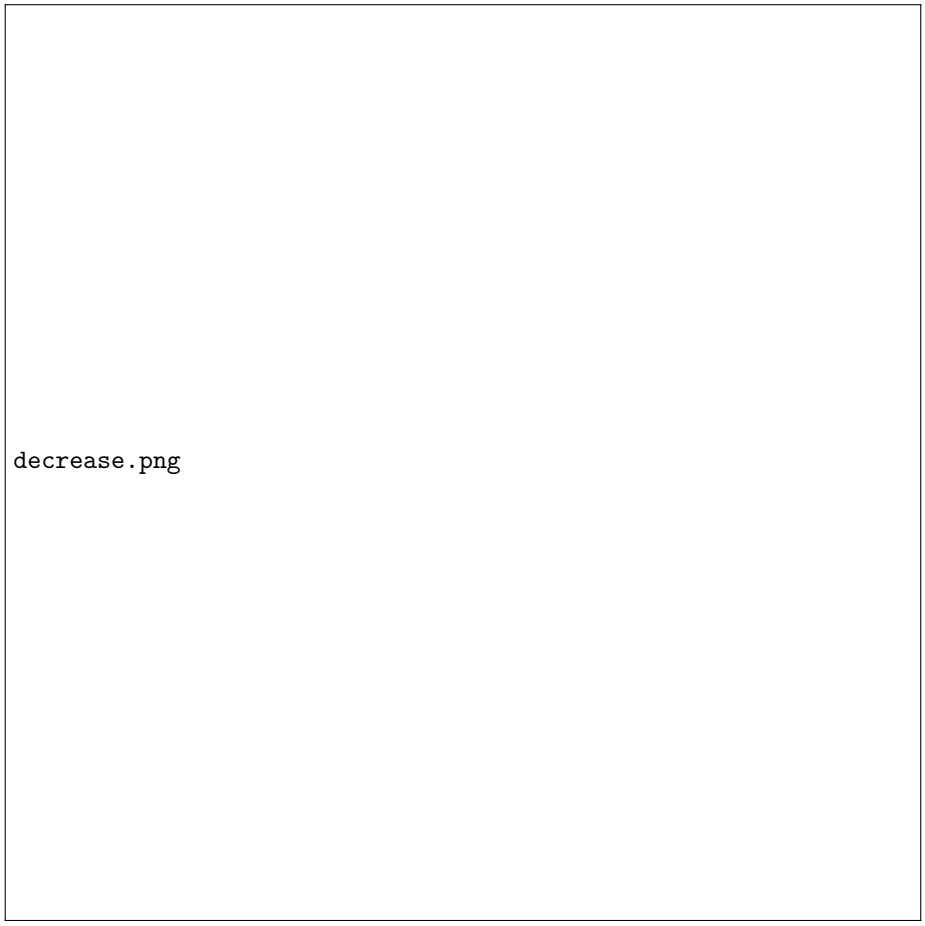

Figure 11: Decreasing step-size$\alpha_t = \frac{10}{t}$

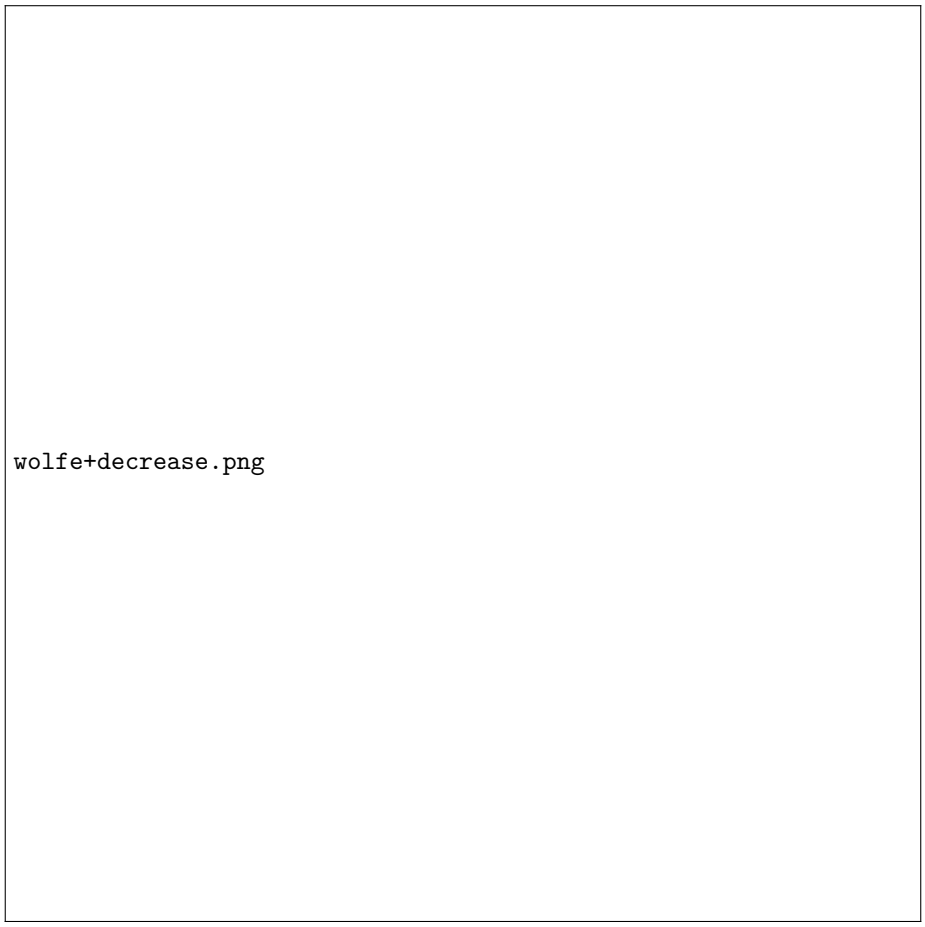

Figure 12: Combined step-size