Mathematics of Reinforcement Learning

Exercise Class 6

Exercise 1 (Bellman optimality equation in the optimal investment problem)**.**
Consider the setting of the optimal investment problem discussed in exercise class 2 task 4. Let $W\colon S \to \mathbb{R}$ be a function defined as

$$\begin{aligned} W(\dagger) &\coloneqq 0, \\ W(T,p,w) &\coloneqq \log(w) \quad \text{and} \\ W(t,p,w) &\coloneqq \max_{a \in A(t,p,w)} \sum_{s' := (t+1,p',w') \in S} p(s' \mid (t,p,w),a) V(s'), \end{aligned}$$

for any $(t,p,w) \in S$ with $t < T$.
Show that $W = V$, where V is the optimal value function and that policies $\pi^* \in \Pi_d$, for which for all $s = (t,p,w) \in S$, with $t < T$ an action $a(s) \in A$ exists with

$$\pi^*(a(s) \mid s) = 1 \quad \text{and} \quad a(s) \in \operatorname{argmax}_{a \in A(s)} \sum_{s' := (t+1,p',w') \in S} p(s' \mid s,a) V(s'),$$

it holds that π^* is optimal for all s.
Hint: Argue analogously to the proof of Theorem 2.9. Start by showing that $W(\dagger) = V(\dagger)$, continue by showing that $W(T,\cdot) = V(T,\cdot)$. Finally, perform a backwards induction, to show that the policy is optimal and that $V = W$. ◇

Exercise 2 (Programming task)**.**
Calculate an optimal policy for the optimal investment problem using the Bellman optimality equation.

◇