Mini-Course 1, Module 3
Value Functions & Bellman Equations

CMPUT 397
Fall 2020

Reminders: Sept 21, 2019

- Announcement sent out about Discussion Sessions
  - Please fill out the Google Form
- Graded Assessment for Course 1, Module 3 (Graded Quiz) due **this Friday**
- Any questions about admin?

Review of Mini-Course 1, Module 3

Video 1: Policies

- All about policies. All about how our agents **select actions**
- Goals:
  - recognize that a policy is a **distribution** over actions for each state
  - describe the similarities and differences between **stochastic** and **deterministic policies**
  - generate valid policies for a given MDP, or Markov Decision Process.

Video 2: Value Functions

- All about value functions, the key data structure of RL
- Goals:
  - describe the roles of the **state-value** and **action-value** functions in reinforcement learning
  - describe the relationship between **value functions** and **policies**
  - create examples of value functions for a given MDP.

Video 3: Bellman Equation Derivation

- Bellman equations: the foundation of many RL algorithms
- Goals:
  - derive the Bellman equation for **state value functions**
  - derive the Bellman equation for **action-value functions**
  - understand how Bellman equations relate **current and future values**.

Video 4: Why Bellman Equations?

- Why are Bellman equations so important in RL
- Goals:
  - use the Bellman equations to **compute** value functions
  - understand how Bellman Equations will allow our algorithms to make updates now, to take into account the future

Video 5: Optimal Policies

- Formalizing our goals: policies that obtains as much reward as possible in the long run
- **Goals**:
  - define an **optimal** policy
  - understand how a policy can be **at least as good** as every other policy in every state
  - Identify an optimal policy for a given MDP.

Video 6: Using Optimal Value Functions to get Optimal Policies

- A hint of how our agents might use value functions to select actions
- **Goals**:
  - understand the connection between the optimal value function and optimal policies
  - **verify** the optimal value function for given MDPs.

On Whiteboard

- Go over expectation form for the Bellman equation
- Revisit a couple of Practice Quiz questions
  - Q5 and Q6 about shifting rewards
  - Q7 about expressing v^* in terms of q^*

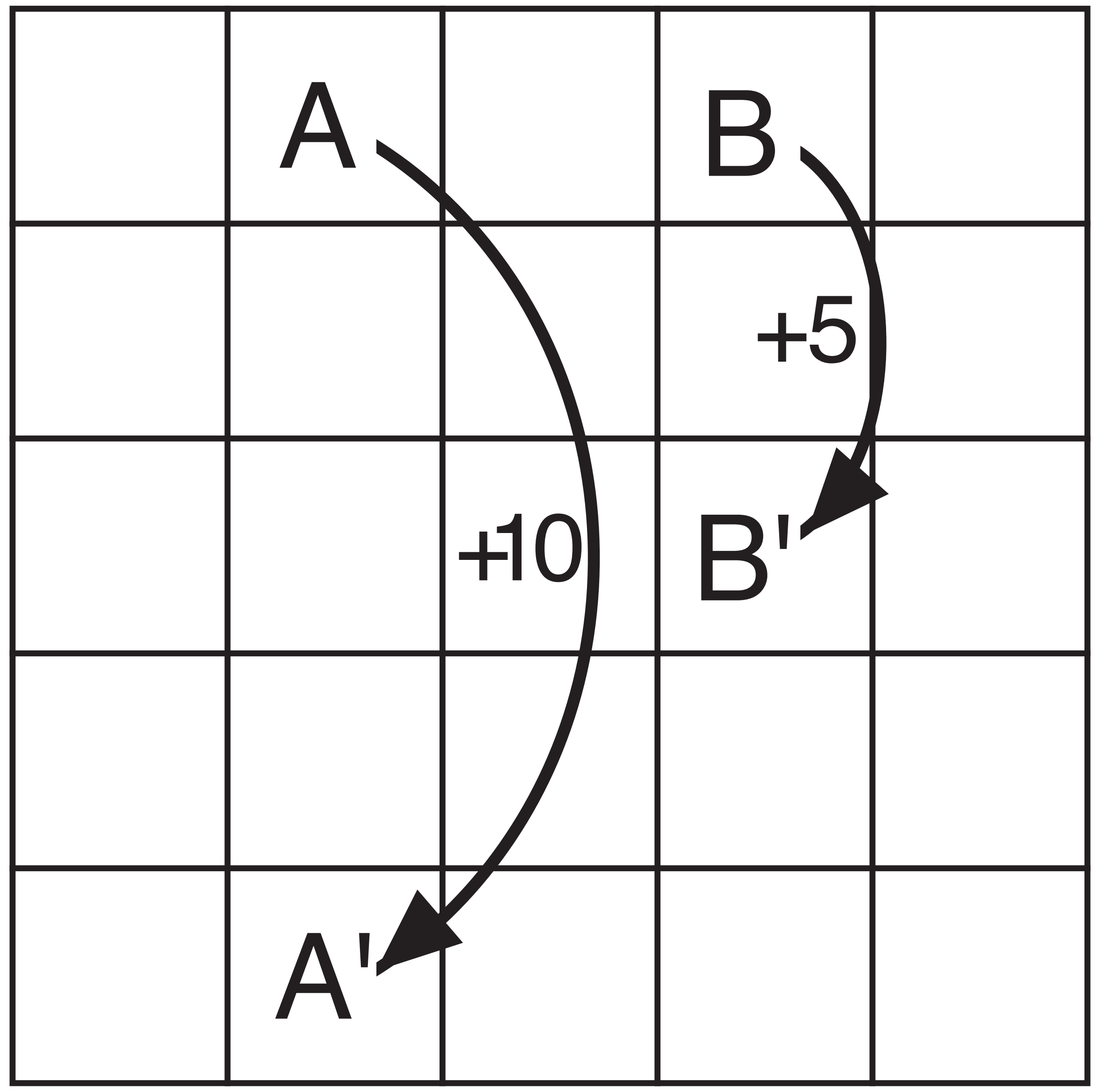
Worksheet Question 1

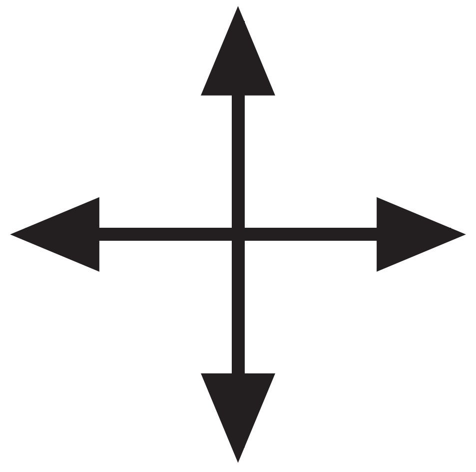
1. Express the action-value function q_π in terms of v_π. The formula will also include p and π.

Practice Question

The Bellman equation (3.10) must hold for each state for the value function v_\pi shown in Figure 3.2. As an example, show numerically that this equation holds for the center state, valued at +0.7, with respect to its four neighboring states, valued at +2.3, +0.4, -0.4, and +0.7. (These numbers are accurate only to one decimal place.). **Harder one:** verify the red state.

$$v_\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s', r \,|\, s, a)\Big[r + \gamma v_\pi(s')\Big], \quad \text{for all } s \in \mathcal{S},$$





Actions

3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

γ = 0.9
π = random
-1 reward on bump