Reinforcement Learning

In-class tutorial:Worked examples [DP, MC, basics of TD]

Subramanian Ramamoorthy
School of Informatics

17 January 2017

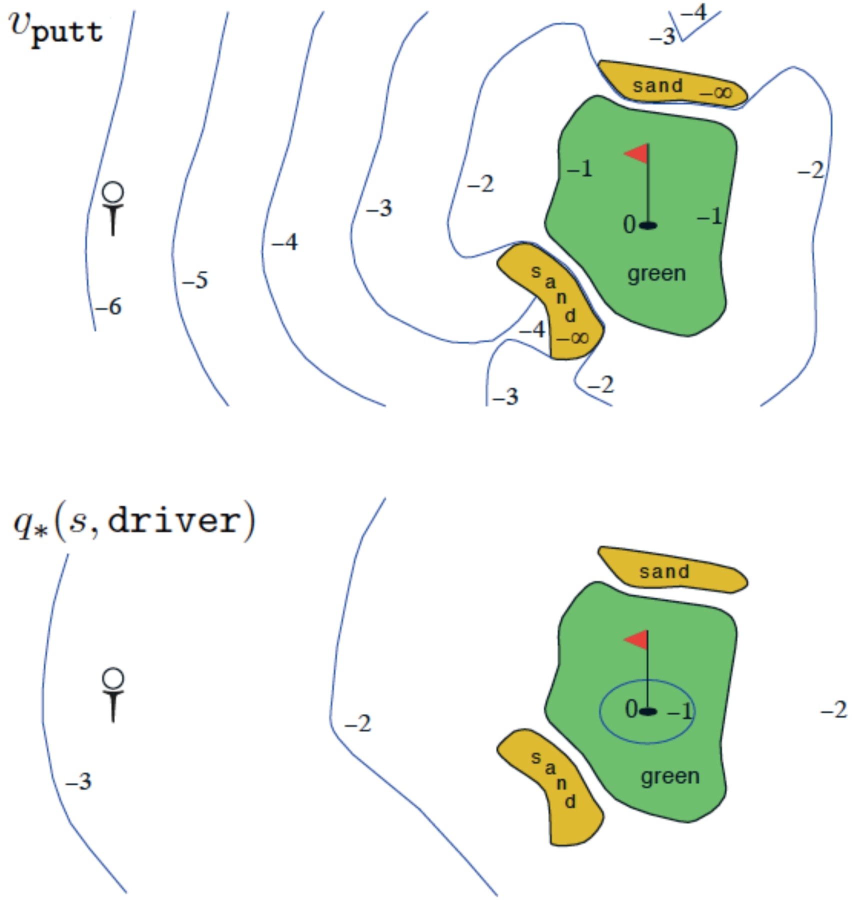
Plan for the Session

- Problems chosen to illustrate concepts covered in earlier lectures

- We will work out problems on the board and take questions to clarify concepts

- These slides provide the outline sketch of the questions to be covered

0. Interpretation of V and Q



Using the task of selecting a club to play the game of golf, discuss the meaning of V and Q

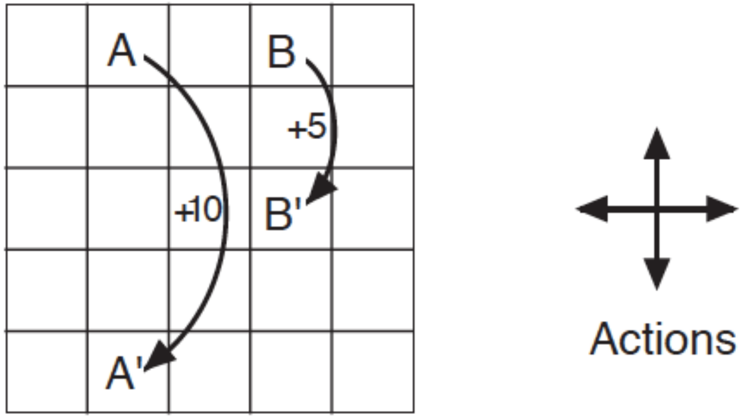
What are:

- States
- Actions
- Rewards

What do you understand by the shape and numbers in this figure?

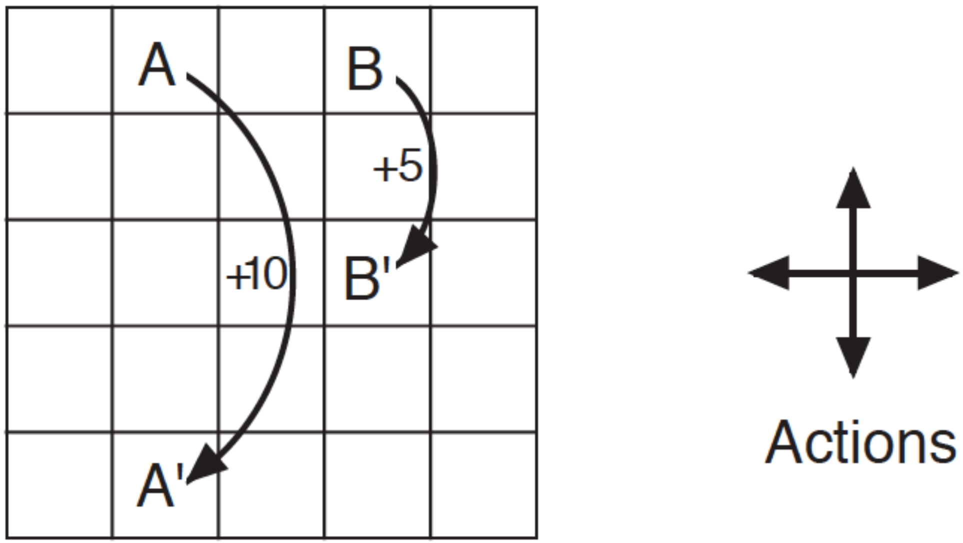
I. Interpretation of V^{π} and π

- Cells = States
- NSEW actions resulting in movement by 1 cell
- Actions taking agent off grid have no effect but incur reward of -1
- All other actions result in a reward of 0
  - except those that move the agent out of the special states A and B.



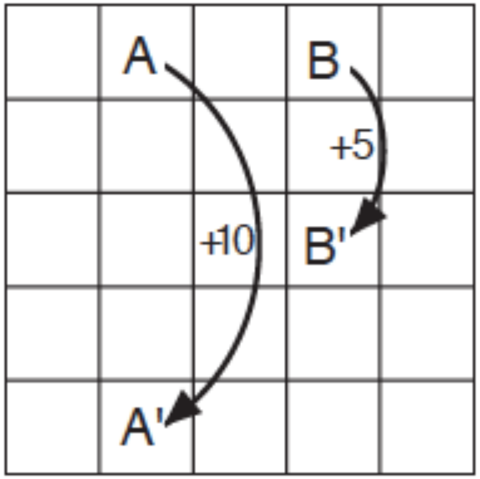
Inspect and interpret V^{π}

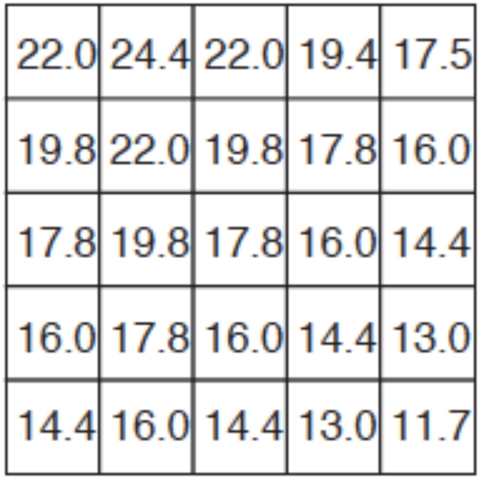
I. Interpretation of V^{π}



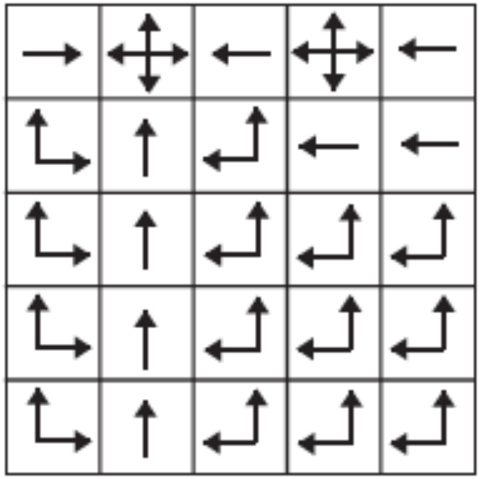
3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

I. Interpretation of V* and π*



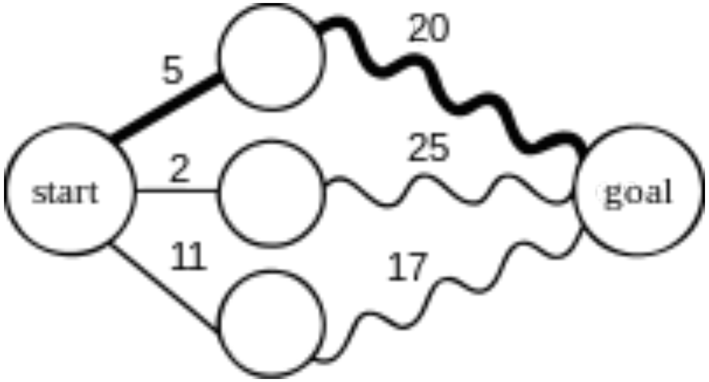


22.0	24.4	22.0	19.4	17.5
19.8	22.0	19.8	17.8	16.0
17.8	19.8	17.8	16.0	14.4
16.0	17.8	16.0	14.4	13.0
14.4	16.0	14.4	13.0	11.7



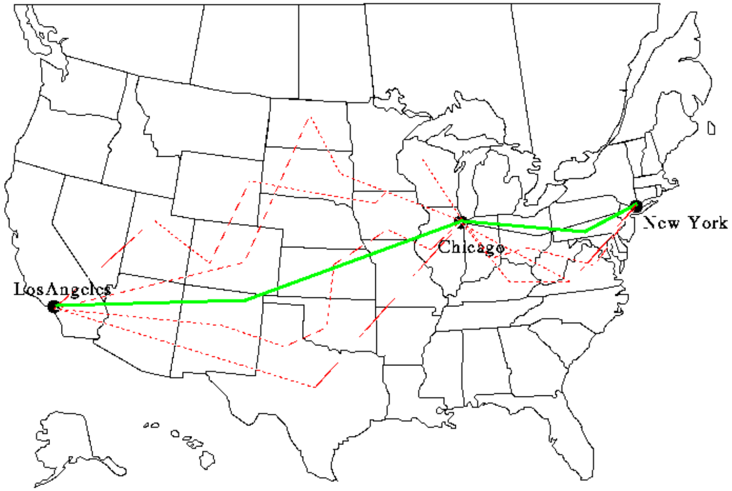
Calculate and show that Bellman's equation holds for centre state – to understand nature of V*

Interpreting V: *Cost-to-go*



Finding the shortest path in a graph using optimal substructure; a straight line indicates a single edge; a wavy line indicates a shortest path between two vertices it connects (other nodes on these paths are not shown); bold line is the overall shortest path from start to goal. [From Wikipedia]

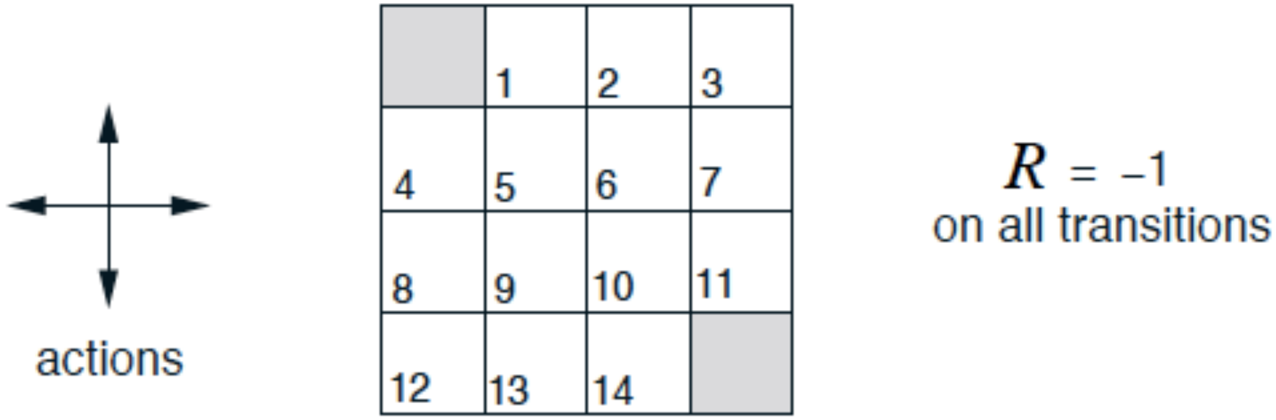
TRIVIAL EXAMPLE OF BELLMAN'S OPTIMALITY PRINCIPLE



Understanding the recursion:
If shortest path from LA to NY must include Chicago, then shortest path from LA to Chicago can be computed separately from last leg.

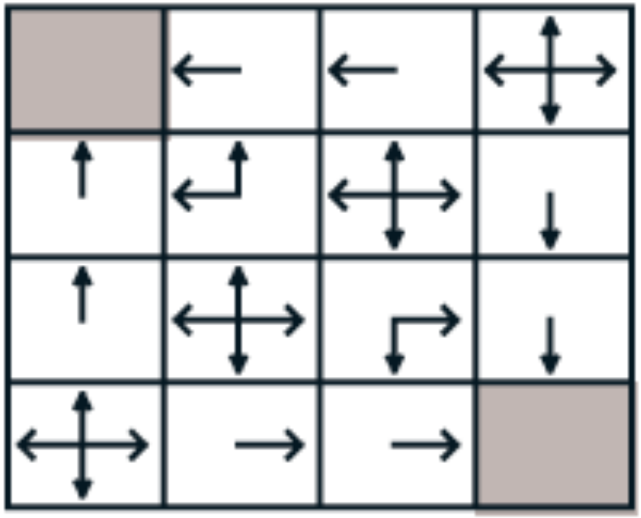
II. Value/ Policy Iteration using Grid World

- Calculate initial steps of Policy Evaluation using a grid world example seen in our earlier lectures



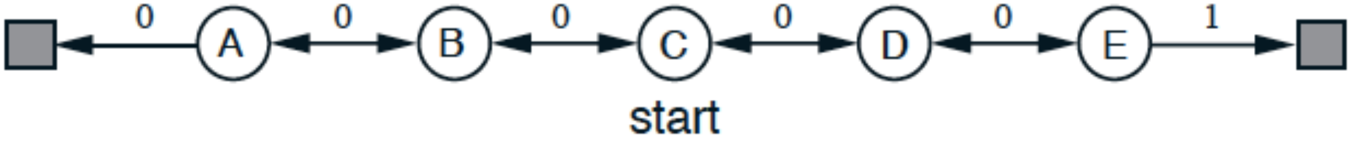
V^{π} and Greedy π at $k = 2$

0.0	-1.7	-2.0	-2.0
-1.7	-2.0	-2.0	-2.0
-2.0	-2.0	-2.0	-1.7
-2.0	-2.0	-1.7	0.0



III. MC Value Evaluation

- Work out some steps of the MC value evaluation process for the 5-state Markov Chain example (for a random walker who goes one step to the left or right with equal probability)



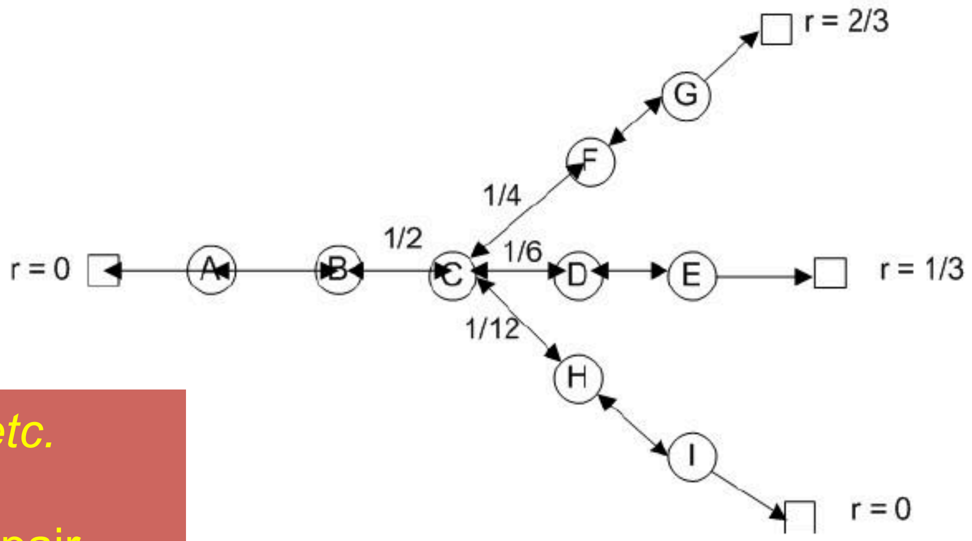
IV. Understanding MC through modified random walk

- The transition probabilities for state C are as shown. For all other states, the transitions are based on a fair coin flip. The square is an absorbing terminal state with reward as shown.

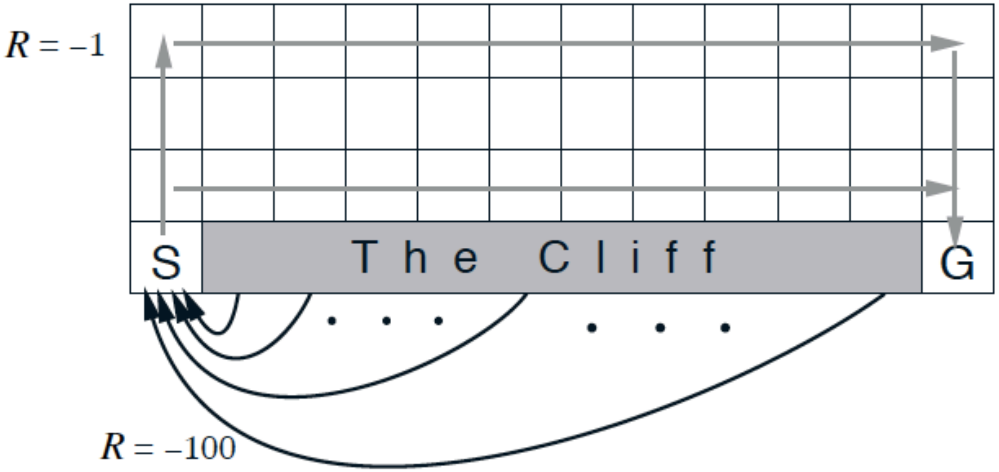
Perform some initial steps of calculation of V^{π} using first-visit MC.

Discuss MC with Exploring Starts, etc.

Exploring starts: Every state-action pair has a non-zero probability of being the starting pair



V. Cliff Walking: TD



Discuss SARSA and Q-learning procedures with respect to this example