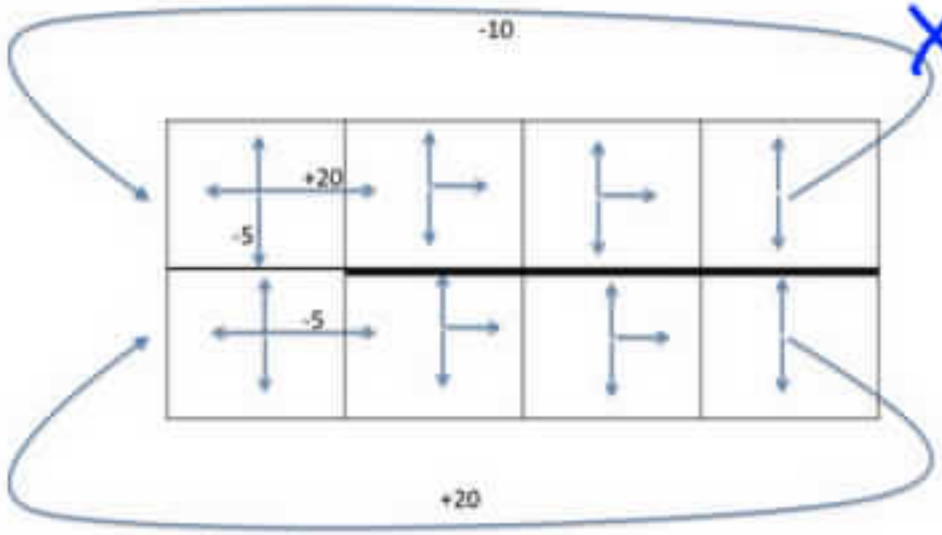
marked. (Note that the marked transitions only give the indicated reward if the action succeeds in moving the agent in that direction.)



$$X = V^{\pi}(s) = R(s) + \gamma \sum_{s'} p(s'|s)\, V^{\pi}(s')$$

$$= \sum_a p(a|s) \sum_{s'} p(s'|s,a)\, R(s,a,s') + \gamma \left(\sum_a p(a|s) \sum_{s'} p(s'|s,a)\, V(s') \right)$$

$$\frac{1}{2}(-2) + \frac{1}{4}\big((.9)(+20) + (.1)(-1)\big) + \frac{1}{4}\big((.9)(-5) + (.1)(-1)\big)$$

(a) MDP (10 pts) Give the MDP for this domain *only for the state transitions starting from each of the states in the top row*, by filling in a state-action-state transition table (showing only the state transitions with non-zero probability). You should refer to each state by its row and column index, so the upper left stat [1,1] and the lower right state is [2,4].

$$+(.85)\left[\frac{1}{2}X + \frac{1}{4}\big((.9)(-11.7) + (.1)X\big) + \frac{1}{4}\big((.9)(-8.5) + (.1)X\big)\right]$$

To get you started, here are the first few lines of the table:

State s	Action a	New state s'	p(s'ls,a)	r(s,a,s')
[1,1]	Up	[1,1]	1.0	-2
[1,1]	Right	[1,1]	0.1	-1
[1,1]	Right	[1,2]	0.9	+20

(b) Value function (15 pts) Suppose the agent follows a randomized policy π (where each available action in any given state has equal probability) and uses