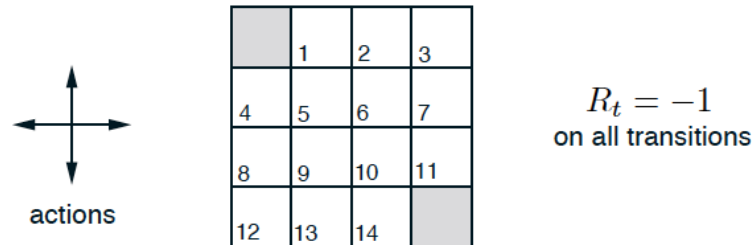


## Homework 2

Due date: 21<sup>st</sup> May, 2020

1. (100 points) **(Matlab Exercise for the gridworld)** Consider the 4×4 gridworld shown below.



The nonterminal states  $\mathcal{S}=\{1, 2, \dots, 14\}$ . There are four actions possible in each state,  $\mathcal{A}=\{\text{up, down, right, left}\}$ , which deterministically cause the corresponding state transitions, except that actions would take the agent off the grid in fact leave the state unchanged. This is an undiscounted ( $\gamma=1$ ), episodic task. The reward is -1 on all transitions until the terminal state is reached. The terminal state is shaded in above figure. Suppose that the agent follows the equiprobable random policy (all actions equally likely). Find the value function for each state of this equiprobable random policy when value converges. (Note: please print the value of each state like the example in the textbook as illustrated in the following. Only print the converged values).

$v_k$  for the  
random policy

$k = 0$

0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0

$k = 1$

0.0	-1.0	-1.0	-1.0
-1.0	-1.0	-1.0	-1.0
-1.0	-1.0	-1.0	-1.0
-1.0	-1.0	-1.0	0.0

$k = 2$

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

$k = 3$

0.0	-2.4	-2.9	-3.0
-2.4	-2.9	-3.0	-2.9
-2.9	-3.0	-2.9	-2.4
-3.0	-2.9	-2.4	0.0

$k = 10$

0.0	-6.1	-8.4	-9.0
-6.1	-7.7	-8.4	-8.4
-8.4	-8.4	-7.7	-6.1
-9.0	-8.4	-6.1	0.0

$k = \infty$

0.0	-14.	-20.	-22.
-14.	-18.	-20.	-20.
-20.	-20.	-18.	-14.
-22.	-20.	-14.	0.0