Ex 02 RL Samuel Pilch, Jakob Gresell

1) a) chess $\mathcal{S}$ = {every possible position}

64 squares × 13 figures (6 black, 6 white, empty) => 2 dim discrete

$\mathcal{A}$ = {every legal move}

16 figures × different moves with each figure => 2 dim discrete

$\mathcal{R}$ = {game ended in win/lose} => 1 dim discrete

b) pick & place $\mathcal{S}$ = {position endeffector, position object, item grabbed}

6 dim cont × 6 dim cont × 1 dim discrete

$\mathcal{A}$ = {change position endeffector, grabbing}

6 dim cont × 1 dim discrete

$\mathcal{R}$ = {MSE of object and goal}

6 dim cont.

c) drone $\mathcal{S}$ = {position, angles and derivatives}

12 dim cont

$\mathcal{A}$ = {voltage of motors}

4 dim cont

$\mathcal{R}$ = {MSE of current position, angles and derivatives to goalstate (0)}

1 dim cont

d) vacuum robot $\mathcal{S}$ = {x, y, φ, on/off}

3 dim cont × 1 dim discrete

$\mathcal{A}$ = {acceleration, turning, switching on/off}

2 dim cont × 1 dim discrete

$\mathcal{R}$ = {sum of vacuumed area + derivative of vacuumed area}

1 dim cont

2, a) We don't need to consider future rewards in the bandit setting since each action and reward are independent of earlier and later actions, so the bandits don't need a state.

b) $v_\pi(s) \overset{\text{slide 31}}{=} \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_\pi(s')\right] \quad \forall s \in S \quad (1)$

$q_\pi(s) \overset{\text{slide 33}}{=} \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_\pi(s')\right]$

Since this right hand site of the equation is also present in equation (1)

$\Rightarrow v_\pi(s) = \sum_a \pi(a|s)\, q_\pi(s)$

c) $v_\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_\pi(s')\right]$

$= \sum_a \pi(a|s) \sum_{s'} \sum_r p(s',r|s,a)\left[r + \gamma v_\pi(s')\right]$

$\overset{\text{slide 21}}{=} \sum_a \pi(a|s) \sum_{s'} p(s'|s,a)\, r(s,a,s') + \sum_r p(s',r|s,a)\, \gamma\, v_\pi(s')$

$= \sum_a \pi(a|s) \sum_{s'} p(s'|s,a)\left(r(s,a,s') + \gamma v_\pi(s')\right)$

3, a) Be $\mathcal{P} = \{\pi_i\}$

then $|\mathcal{P}| = |A|^{|S|}$

b) $v_\pi = R + \gamma P_\pi v_\pi$

$\Leftrightarrow v_\pi - \gamma P_\pi v_\pi = R$

$\Leftrightarrow v_\pi (I - \gamma P_\pi) = R$

$\Leftrightarrow v_\pi = (I - \gamma P_\pi)^{-1} R$

3, c) $V^* = \begin{bmatrix} 0{,}498 & 0{,}831 & 1{,}311 \\ 0{,}536 & 0{,}977 & 2{,}295 \\ 0{,}306 & 0 & 5 \end{bmatrix}$

There exist 2 optimal policies

→	→	→
↑	↑	→
←	–	–

d) The method doesn't work well, since computation time gets very large. This solution method need small discrete state and action spaces.