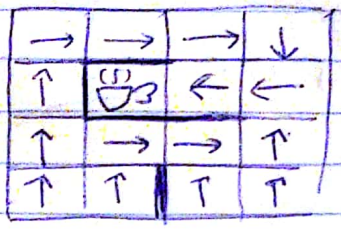
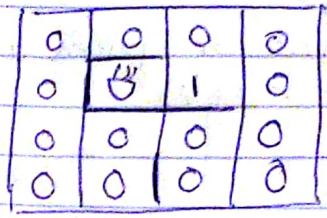
(a) No of states : $4 \times 4 = 16$
No of actions : 3

(b)



After one synchronous iteration



Working backwards from goal

(3,3) $1 + 0.9(0) = 1$

(3,4) $0 + 0.9(1) = 0.9$

(4,4) $0 + 0.9(0.9) = 0.81$

(4,3) $0 + 0.9(1) = 0.9$

(4,2) $0 + 0.9(0.9) = 0.81$

(4,1) $0 + 0.9(0.81) = 0.729$

(3,1) $0 + 0.9(0.729) = 0.6561$

(2,1) $0 + 0.9(0.6561) = 0.5904$

(1,1) $0 + 0.9(0.5904) = 0.5314$

(2,4) $0 + 0.9(0.9) = 0.81$

(1,4) $0 + 0.9(0.81) = 0.729$

(2,3) $0 + 0.9(0.81) = 0.729$

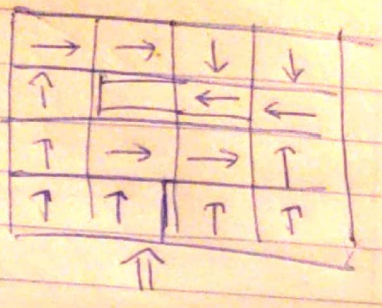
(1,3) $0 + 0.9(0.729) = 0.6561$

(2,2) $0 + 0.9(0.729) = 0.6561$

(1,2) $0 + 0.9(0.6561) = 0.5904$

(c) continued...

0.729	0.81	0.9	0.81
0.6561		1	0.9
0.5904	0.6561	0.729	0.81
0.5314	0.5904	0.6561	0.729

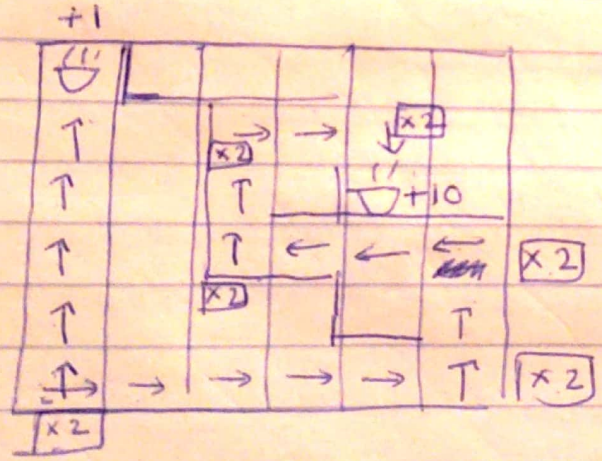


optimal policy

⇐ corresponding value function for each state

(d) The policy in (b) is not the optimum policy
Because it is not equivalent to the policy in (c)
The value function in (3, 4) has not converged yet

(e)



Consider state at start point :

For small cup

$Q_1 = y^4$

For large cup

$Q_2 = y^{20} \times 10$

To choose large coffee over small

$10y^{20} > y^4 \Rightarrow 10y^{16} > 1 \Rightarrow y^{16} > 0.1$

$\gamma > 0.86596$

The agent will choose the big cup over the small one if $0.86596 < \gamma < 1$

It will choose the small cup if $0 < \gamma < 0.86596$

f) $\gamma = 1 \quad r < 0$

Consider state at start point

For small cup: (working backward for optimal policy

* calculating value function $r + \gamma(v)$

(5, 1) *	$r + 1 + 1(0) = r + 1$
(4, 1)	$r + 1(r + 1) = 2r + 1$
(3, 1)	$r + 1(2r + 1) = 3r + 1$
(2, 1)	$r + 1(3r + 1) = 4r + 1$
(1, 1)	$r + 1(4r + 1) = 5r + 1 \quad Q_1 = 5r + 1$

Similarly for small cup

$Q_2 = 21r + 10$ (21 steps)

Agent chooses large cup of coffee if

$21r + 10 > 5r + 1$

$16r > -9$

$r > -\frac{9}{16}$

$r > -0.5625$

Agents chooses big cup over small if

$-0.5625 < r < 0.$

It chooses small cup

$r < -0.5625$