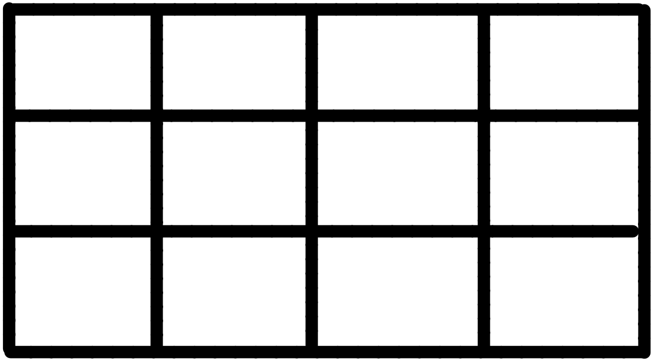
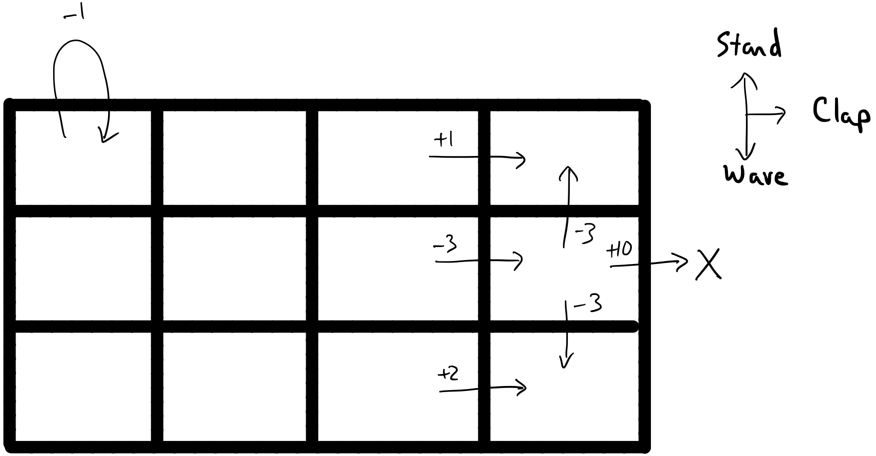


Stand

Clap

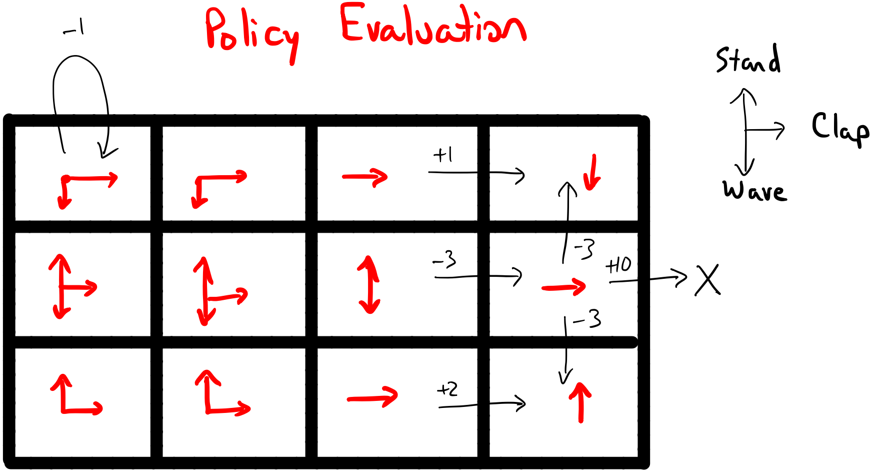
Wave

(RL: no model)

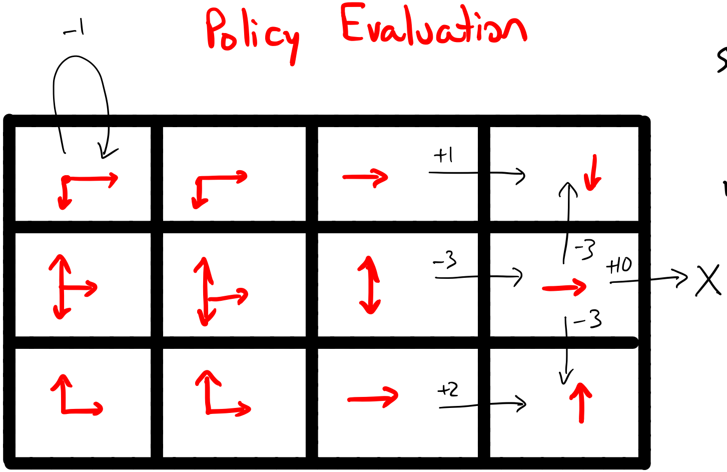


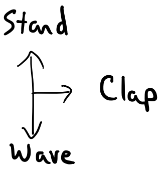
(DP: Known model)

Policy Evaluation



Policy Evaluation





0	0	0	0
0	0	0	0
0	0	0	0

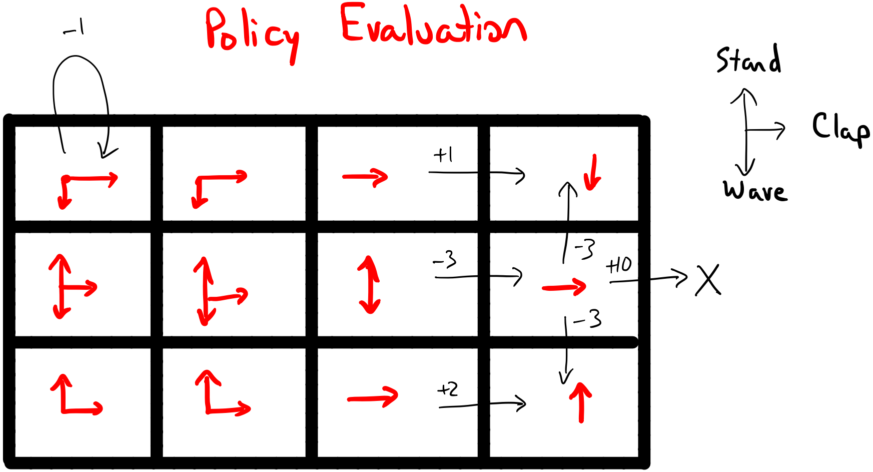
$\boxed{0}$

required

$\forall s \quad v_0(s) = 0$

$v_{k+1}(s) =$ (Eq. 4.5)

Policy Evaluation

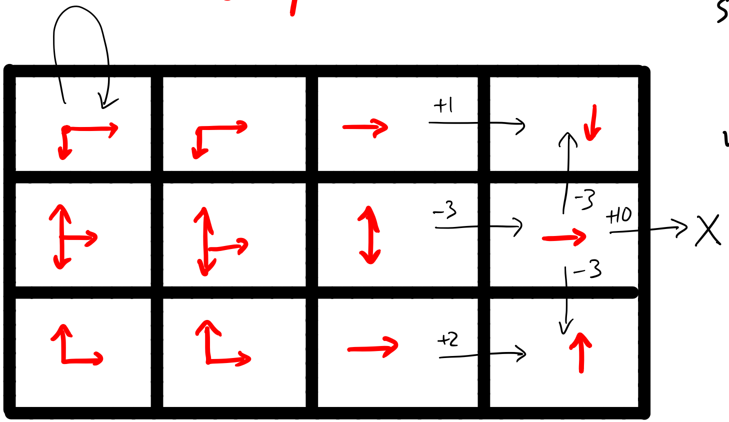


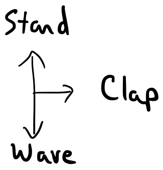
0	0	0	0
0	0	0	0
0	0	0	0

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \times ?$$

Policy Evaluation



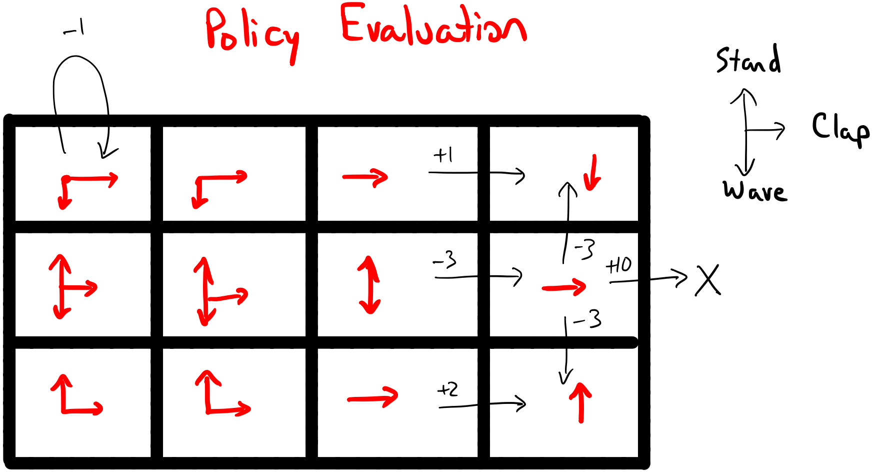


0	0	0	0
0	0	0	0
0	0	0	0

$$\forall s \quad v_0(s)=0$$

$$v_{k+1}(s)=\sum_{a}\pi(a|s)\sum_{s',r}p(s',r|s,a)\times ?$$

Policy Evaluation

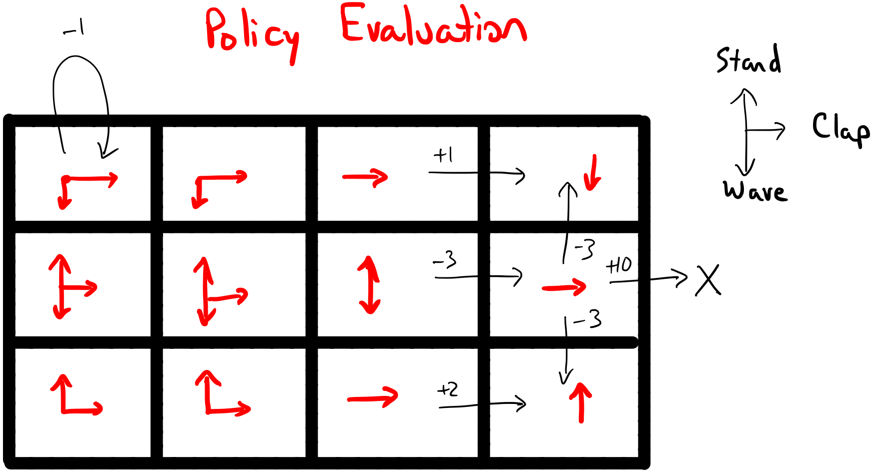


0	0	0	0
0	0	0	0
0	0	0	0

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[r + \gamma v_k(s')]$$

Policy Evaluation

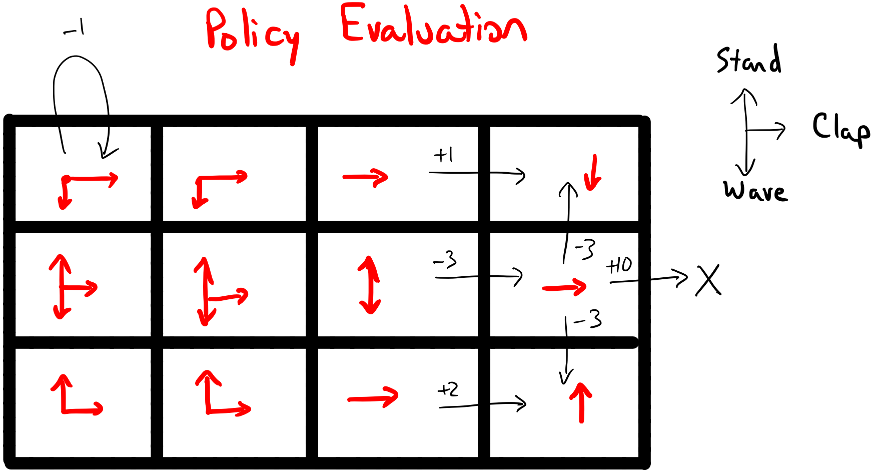


0,0	0,0	0,1	0,0
0,0	0,0	0,0	0,10
0,0	0,0	0,2	0,0

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_k(s')\right]$$

Policy Evaluation

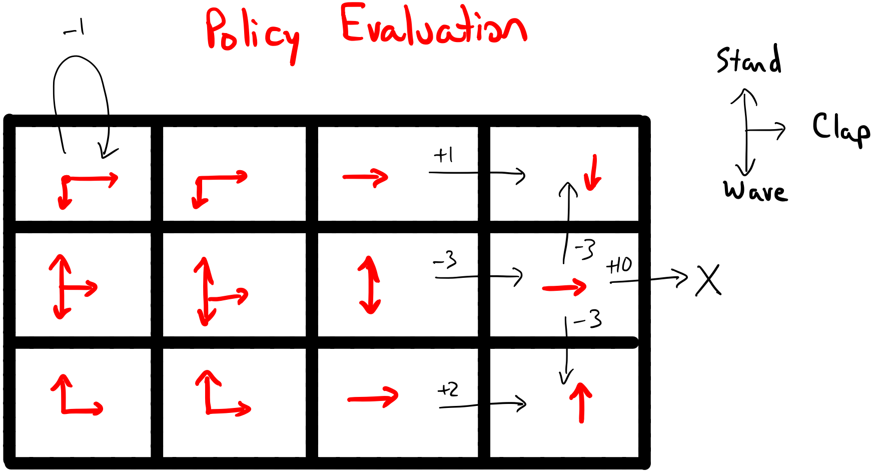


0,0,0	0,0,1,.5	0,1,1,11	0,0,10
0,0,0	0,0,0	0,0,1.5 ?	0,10,10,
0,0,0	0,0,1,1	0,2,2,12	0,0,10

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_k(s')\right]$$

Policy Evaluation

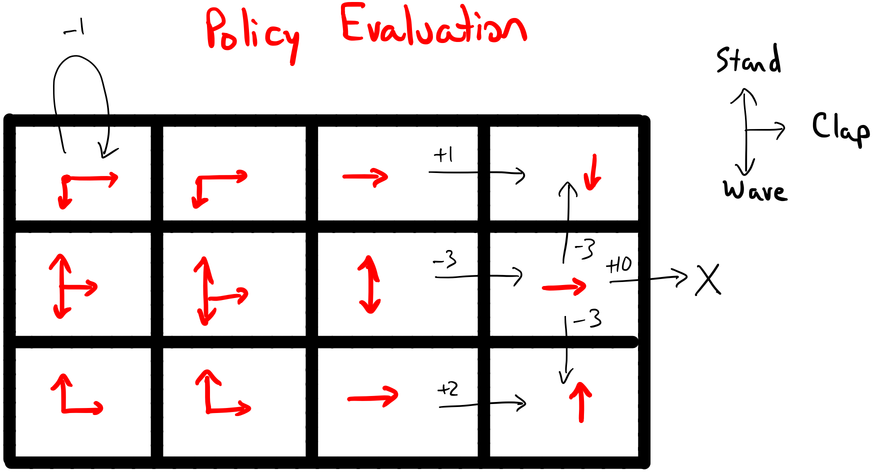


0,0,0	0,0,1,.5 ?	0,1,1,11 11	0,0,10 10
0,0,0	0,0,0	0,0,1.5 11.5	0,10,10, 10
0,0,0	0,0,1,1	0,2,2,12 12	0,0,10 10

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_k(s')\right]$$

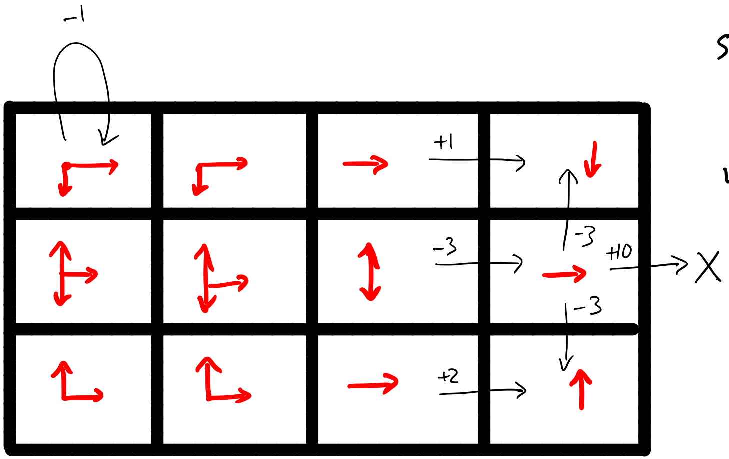
Policy Evaluation

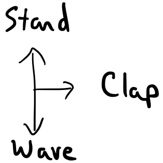


0,0,0 11 3/8	0,0,1,.5 11 1/4	0,1,1,1 11	0,0,10 10
0,0,0 11 1/2	0,0,0 11 1/2	0,0,1.5 11.5	0,10,10, 10
0,0,0 11 3/8	0,0,1,1 11 3/4	0,2,2,12 12	0,0,10 10

$$\forall s \quad v_0(s) = 0$$

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\left[r + \gamma v_k(s')\right]$$



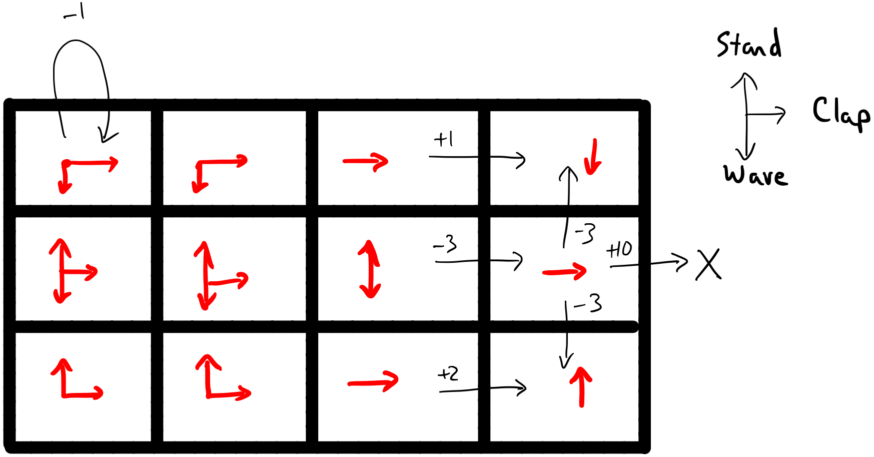


Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

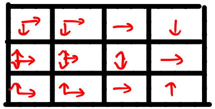
		?	
			?
		?	

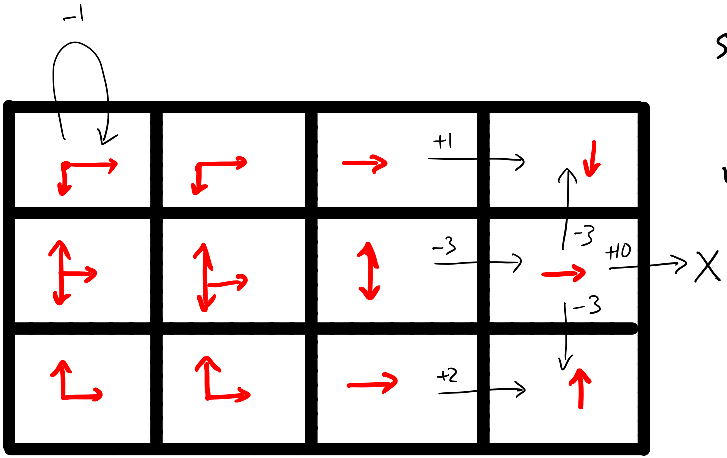


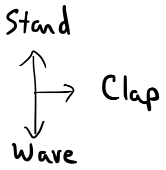
Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

0	0	1	0
0	0	0	10
0	0	2	0







Policy Improvement

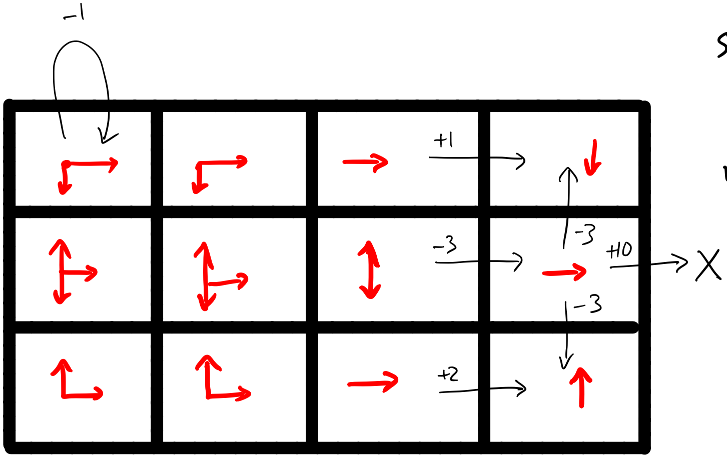
0	0	0	0
0	0	0	0
0	0	0	0

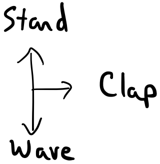
0	0	1	0
0	0	0	10
0	0	2	0

		?	

↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

	→		
		→	
	→		





Policy Improvement

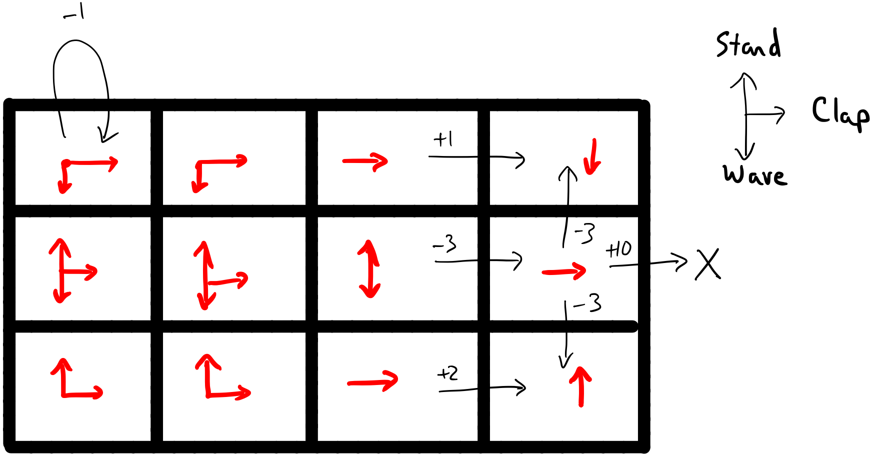
0	0	0	0
0	0	0	0
0	0	0	0

↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

	→		
		→	
	→		

0	1	1	10
0	0	7	10
0	2	2	10



Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

0	0	1	0
0	0	0	10
0	0	2	0

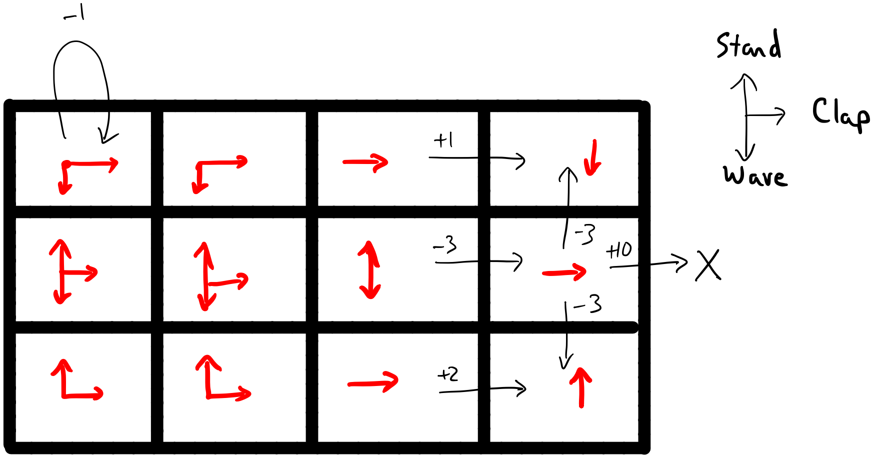
0	1	1	10
0	0	7	10
0	2	2	10

		?	
	?		
		?	

↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

	→		
		→	
	→		

→	→	→	↓
↕→	→	→	→
→	→	→	↑



Policy

0	0	0	0
0	0	0	0
0	0	0	0

0	0	1	0
0	0	0	10
0	0	2	0

0	1	1	10
0	0	7	10
0	2	2	10

1	1	11	10
0	7	7	10
2	2	12	10

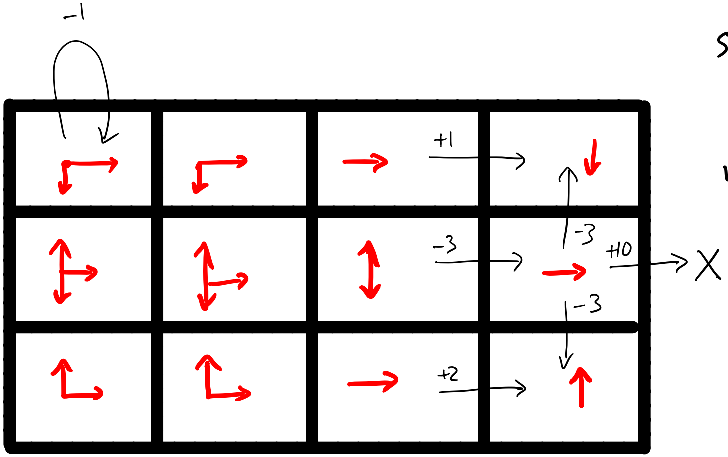
Improvement

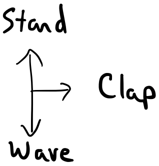
↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

	→		
		→	
	→		

→	→	→	↓
↕→	→	→	→
→	→	→	↑

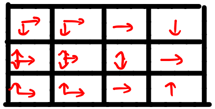
		?	



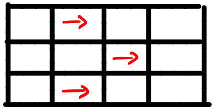


Policy Improvement

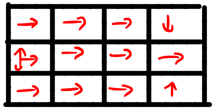
0	0	0	0
0	0	0	0
0	0	0	0



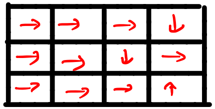
0	0	1	0
0	0	0	10
0	0	2	0



0	1	1	10
0	0	7	10
0	2	2	10

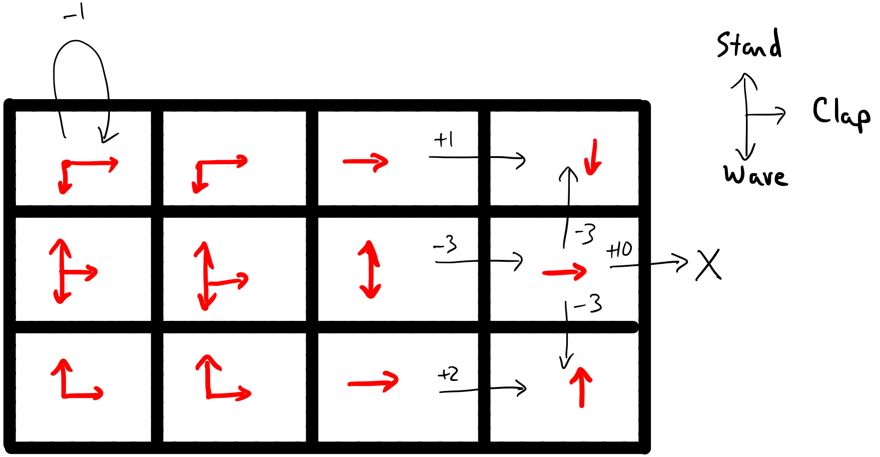


1	1	11	10
0	7	7	10
2	2	12	10



		?	
	?		

Policy Improvement



0	0	0	0
0	0	0	0
0	0	0	0

↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

	→		
		→	
	→		

0	1	1	10
0	0	7	10
0	2	2	10

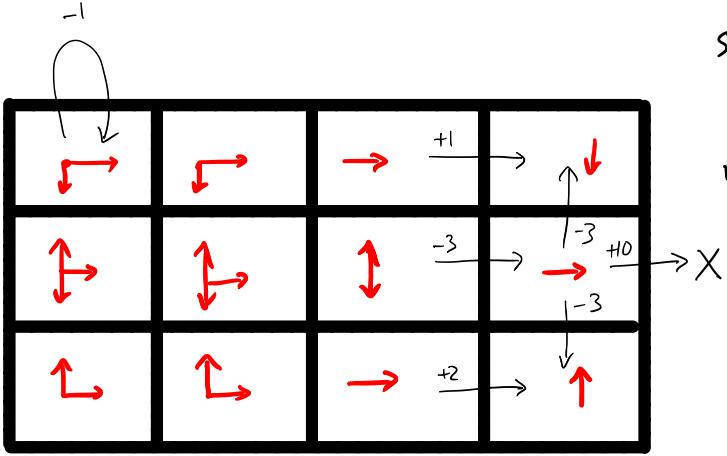
→	→	→	↓
↕→	→	→	→
→	→	→	↑

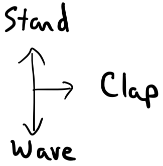
1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	→	↓	→
→	→	→	↑

1	11	11	10
7	7	12	10
2	12	12	10

		?	





Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

	→		
		→	
	→		

0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↕→	→	→	→
→	→	→	↑

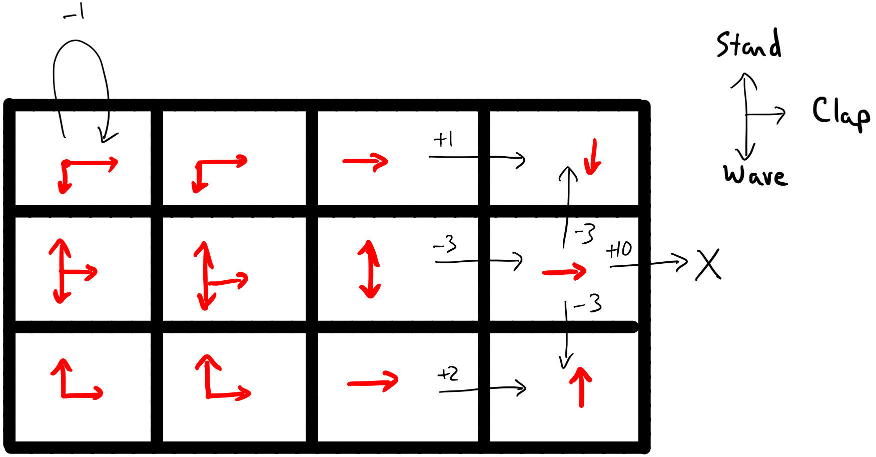
1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	→	↓	→
→	→	→	↑

1	11	11	10
7	7	12	10
2	12	12	10

→	→	↓	↓
→	↓→	↓	→
→	→	↑→	↑

$= \pi^*$



Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

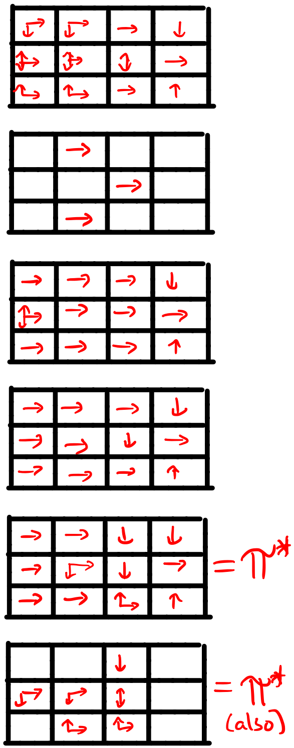
0	0	1	0
0	0	0	10
0	0	2	0

0	1	1	10
0	0	7	10
0	2	2	10

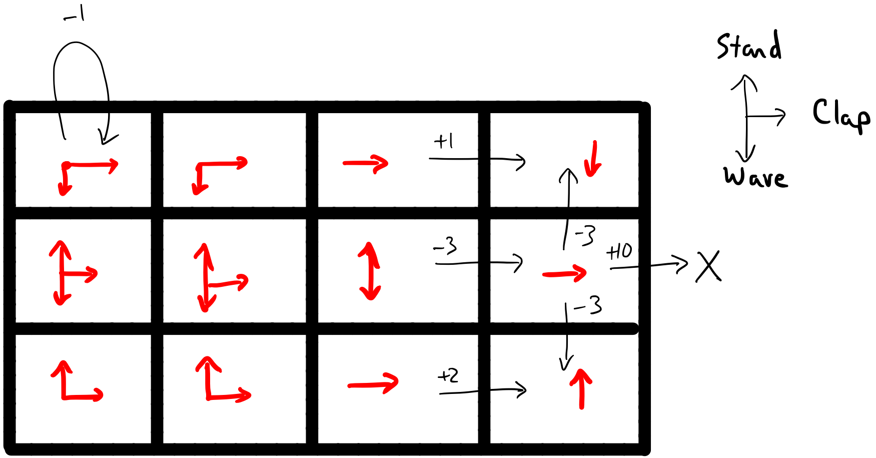
1	1	11	10
0	7	7	10
2	2	12	10

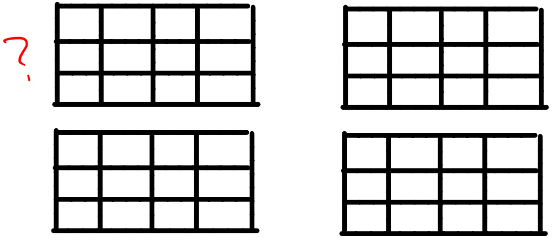
1	11	11	10
7	7	12	10
2	12	12	10

11	11	12	10
7	12	12	10
12	12	12	10



Value Iteration





Policy Iteration

Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

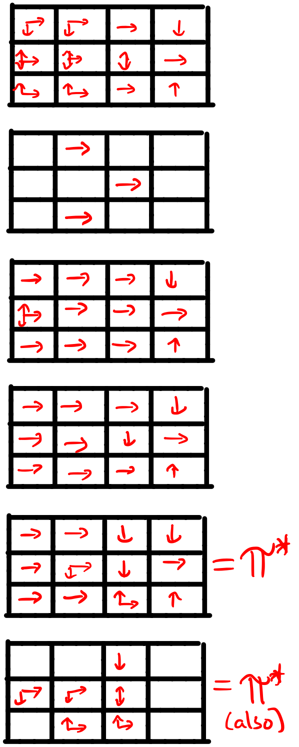
0	0	1	0
0	0	0	10
0	0	2	0

0	1	1	10
0	0	7	10
0	2	2	10

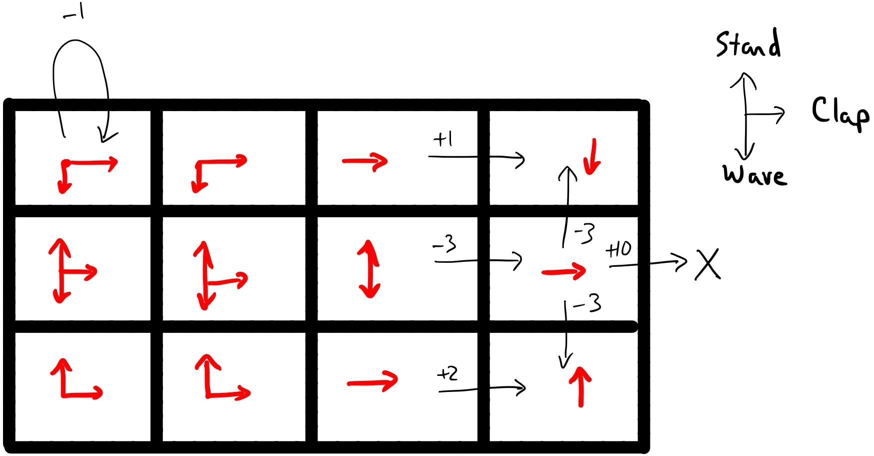
1	1	11	10
0	7	7	10
2	2	12	10

1	11	11	10
7	7	12	10
2	12	12	10

11	11	12	10
7	12	12	10
12	12	12	10



Value Iteration



11 3/8	11 1/4	11	10
11 1/2	11 1/2	11 1/2	10
11 3/8	11 3/4	12	10

?

?

Policy Iteration

Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

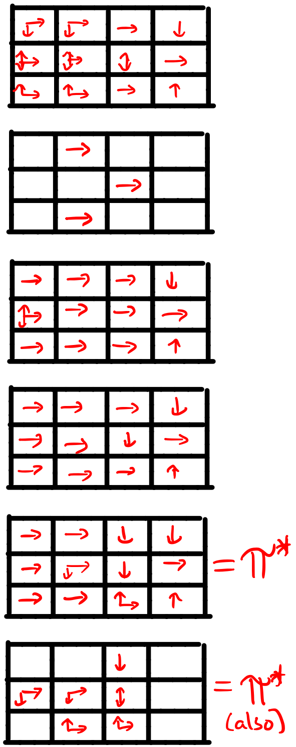
0	0	1	0
0	0	0	10
0	0	2	0

0	1	1	10
0	0	7	10
0	2	2	10

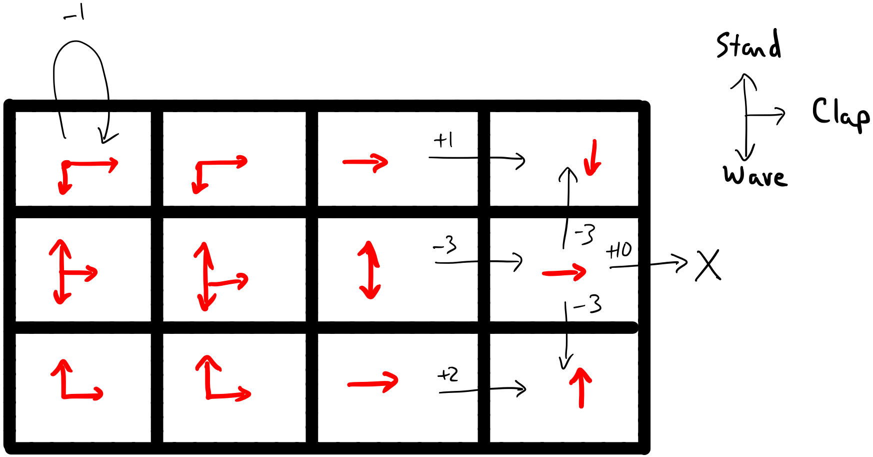
1	1	11	10
0	7	7	10
2	2	12	10

1	11	11	10
7	7	12	10
2	12	12	10

11	11	12	10
7	12	12	10
12	12	12	10



Value Iteration



11 3/8	11 1/4	11	10
11 1/2	11 1/2	11 1/2	10
11 3/8	11 3/4	12	10

↓	↓	↓	↓
→	↓	↓	→
→	→	→	↑

12	12	12	10
12	12	12	10
12	12	12	10

Policy Iteration

Policy Improvement

0	0	0	0
0	0	0	0
0	0	0	0

↓→	↓→	→	↓
↕→	↕→	↕	→
↑→	↑→	→	↑

0	0	1	0
0	0	0	10
0	0	2	0

	→		
		→	
	→		

0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↕→	→	→	→
→	→	→	↑

1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	→	↓	→
→	→	→	↑

1	11	11	10
7	7	12	10
2	12	12	10

→	→	↓	↓
→	↓→	↓	→
→	→	↑→	↑

$= \pi^*$

11	11	12	10
7	12	12	10
12	12	12	10

		↓	
↓→	↓→	↕	
	↑→	↑→	

$= \pi^*$
(also)

Value Iteration