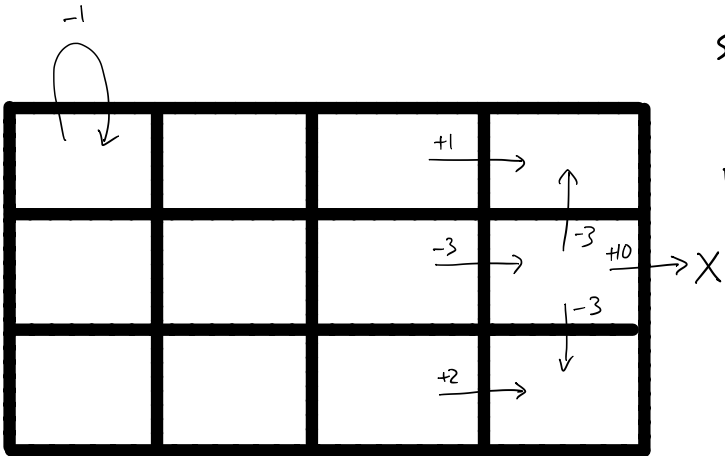


Stand

Clap

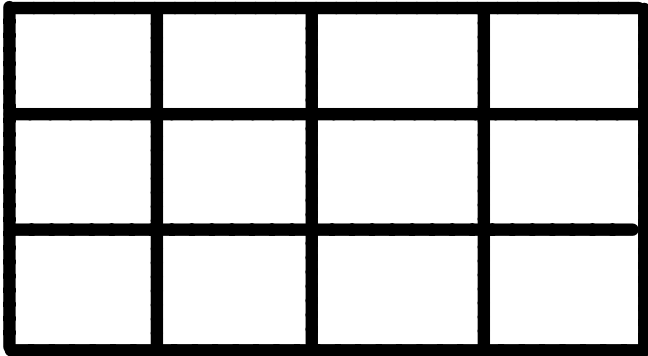
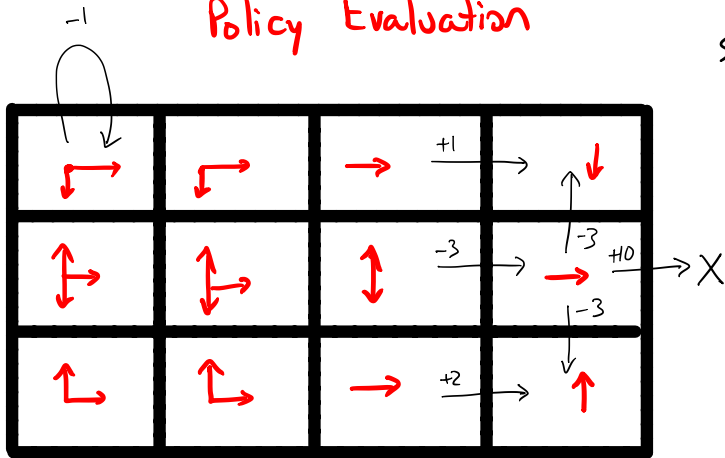
Wave

(RL: no model)

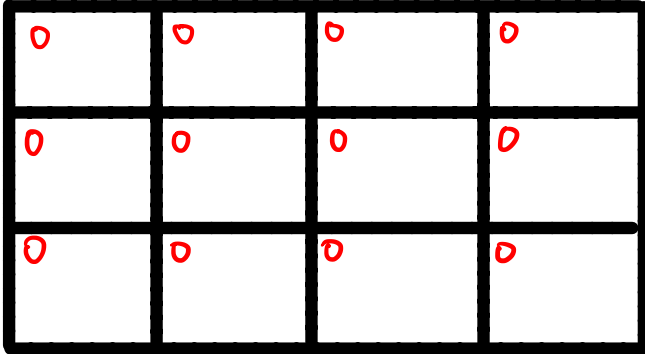
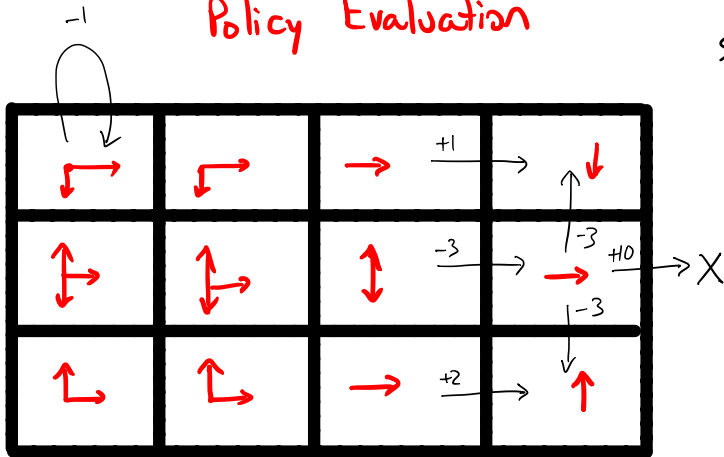


(DP: known model)

Policy Evaluation



Policy Evaluation



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

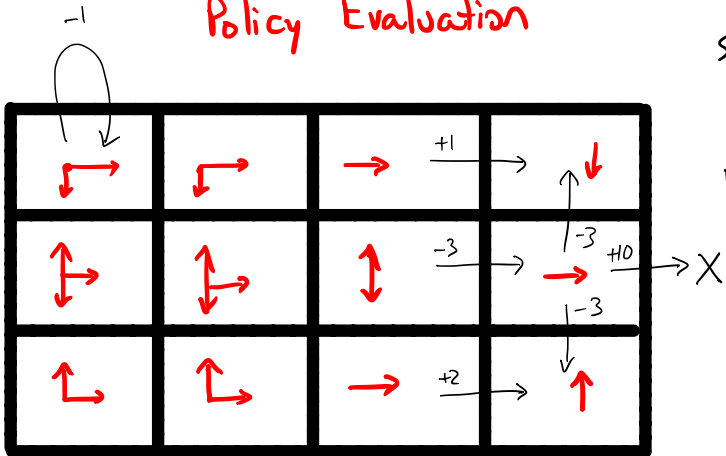
$$v_{k+1}(s) =$$

(Eq. 4.5)



required

Policy Evaluation



Stand
 ↑
 ↓
 Wave

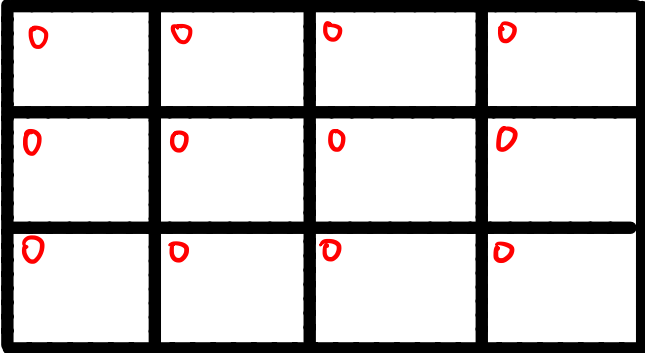
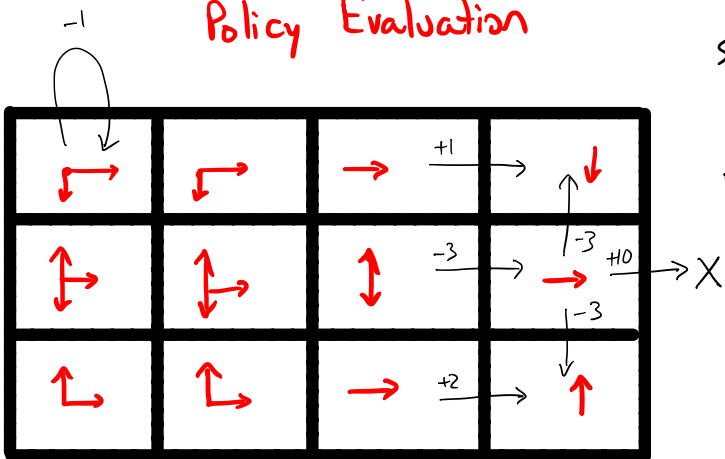
Clap
 →

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

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

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

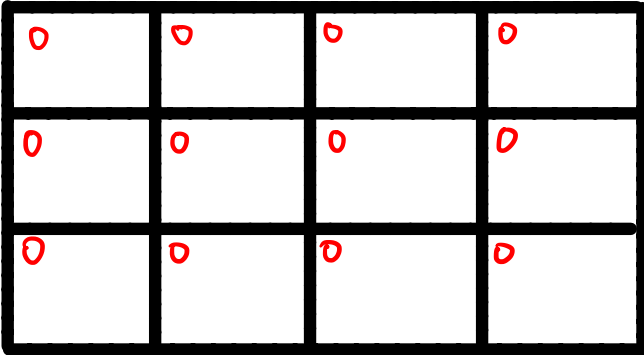
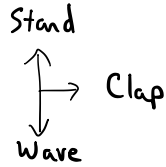
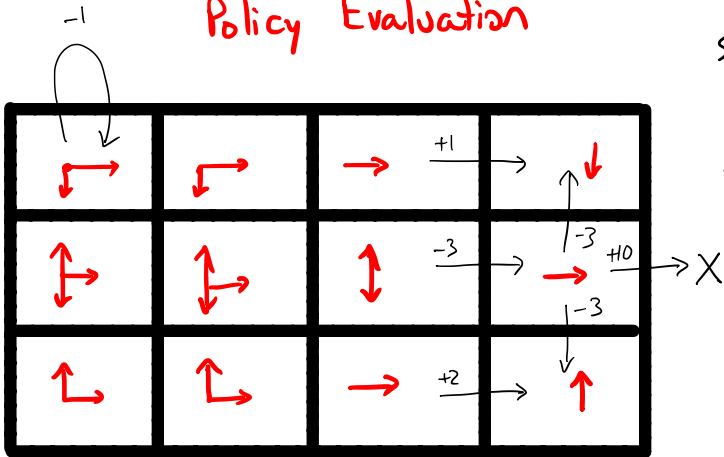
Policy Evaluation



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

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

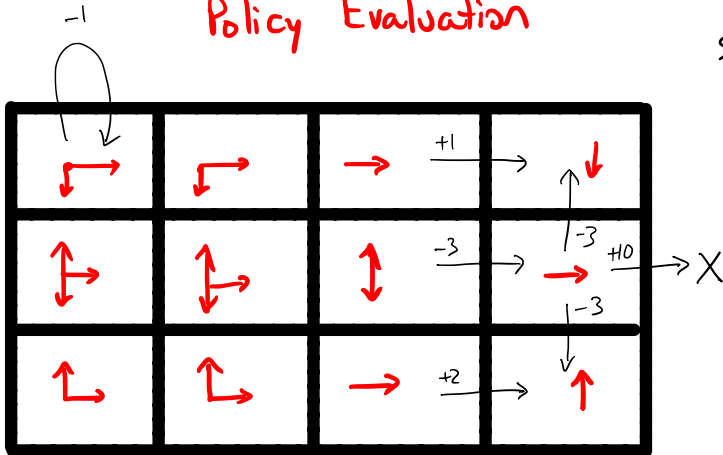
Policy Evaluation



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

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

Policy Evaluation

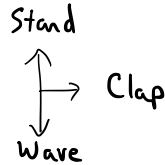
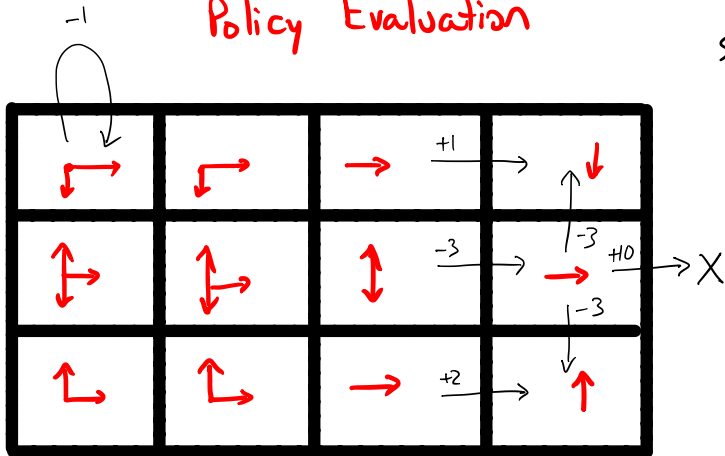


0,0	0,0	0,1	0,0
0,0	0,0	0,0	0,10
0,0	0,0	0,2	0,0

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

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

Policy Evaluation

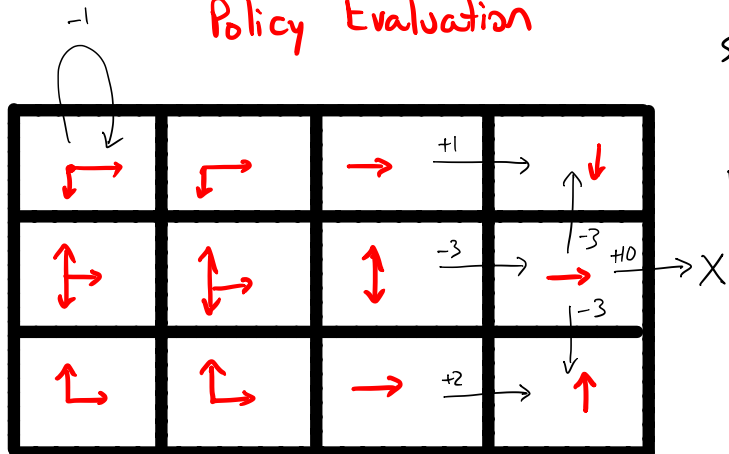


0,0,0	0,0,1,5	0,1,1,11	0,0,10
0,0,0	0,0,0	0,0,15 ?	0,10,10
0,0,0	0,0,1,1	0,2,2,12	0,0,10

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

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

Policy Evaluation



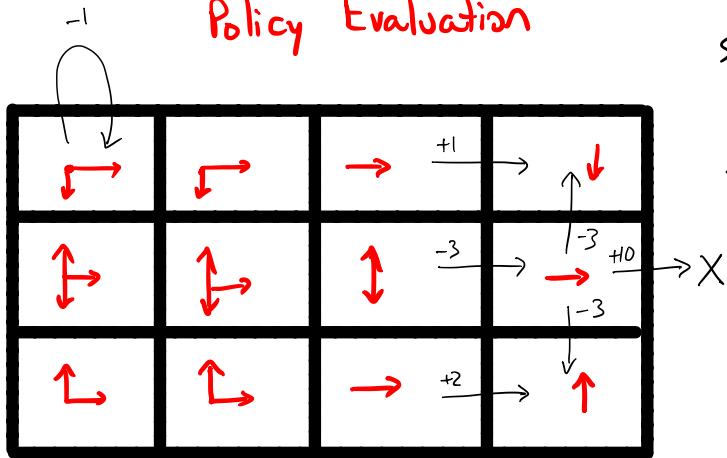
Stand
Wave
Clap

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

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

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

Policy Evaluation

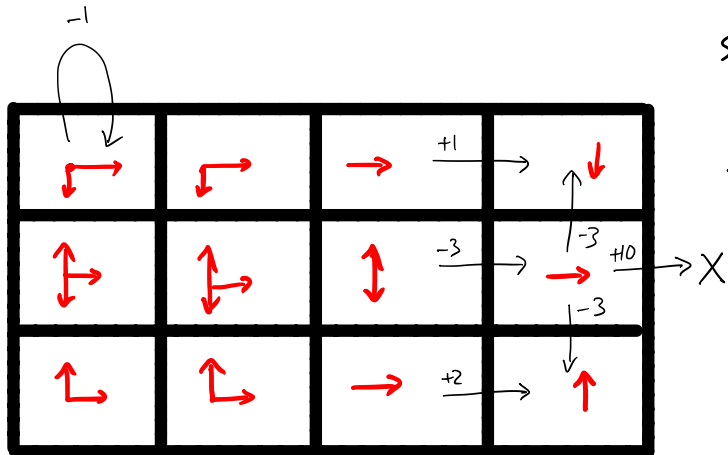


Stand
 ↑
 Clap
 ↓
 Wave

0,0,0 3/8	0,0,1,5 1/4	0,1,1,11 	0,0,10 0
0,0,0 1/2	0,0,0 1/2	0,0,15 5	0,10,10, 0
0,0,0 3/8	0,0,1,1 3/4	0,2,2,12 2	0,0,10 0

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

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



Stand
 Wave
 Clap

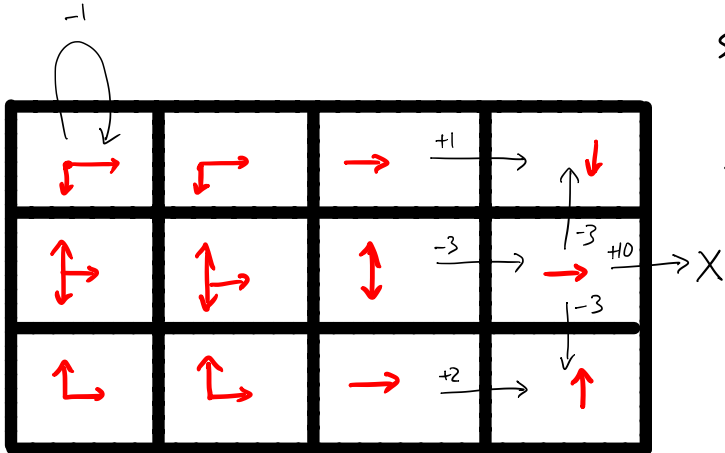
Policy

Improvement

o	o	o	o
o	o	o	o
o	o	o	o

↘	↘	→	↓
↔	↔	↓	→
↺	↺	→	↑

		?	
			?
		?	



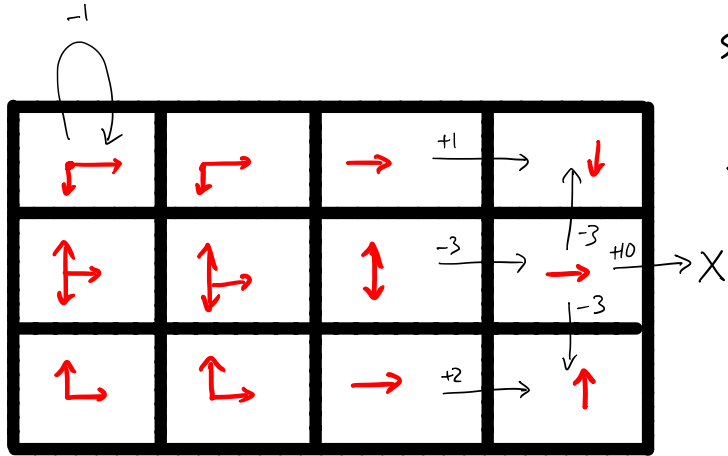
Stand
 ↑
 Clap
 ↓
 Wave

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



Stand
Wave
Clap

Policy Improvement

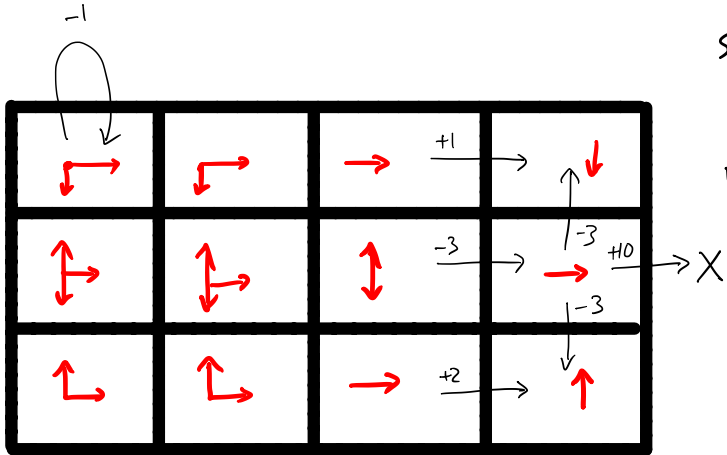
0	0	0	0
0	0	0	0
0	0	0	0

↖	↘	→	↓
↔	↔	↓	→
↖	↖	→	↑

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

	→		
		→	
	→		

		?	



Policy

Improvement

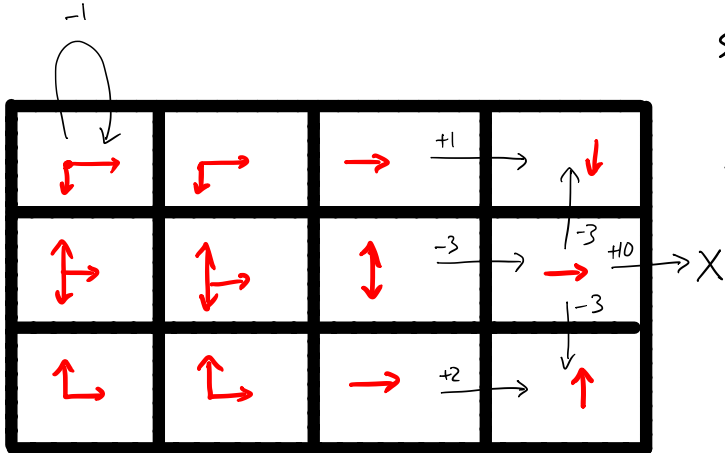
0	0	0	0
0	0	0	0
0	0	0	0

↖	↘	→	↓
↔	↔	↕	→
↻	↻	→	↑

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

	→		
		→	
	→		

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



Stand
Wave
Clap

Policy Improvement

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

↗	↘	→	↓
↔	↔	↓	→
↖	↖	→	↑

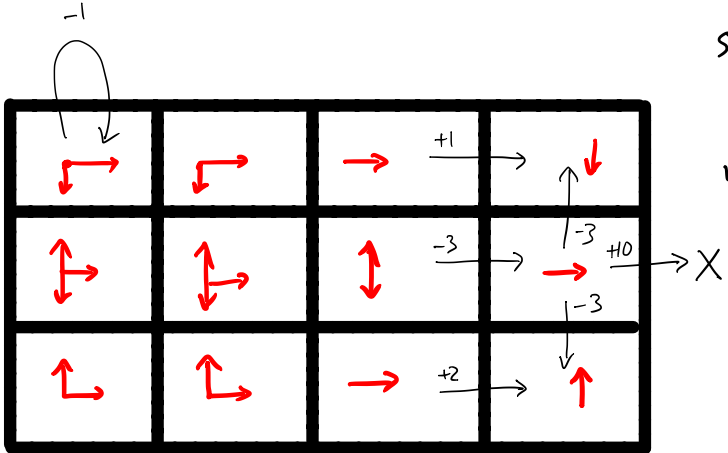
0	0	1	0
0	0	0	10
0	0	2	0

	→		
		→	
	→		

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

→	→	→	↓
↔	→	→	→
→	→	→	↑

		?	
	?		
		?	



Stand
 ↑
 Clap →
 ↓
 Wave

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

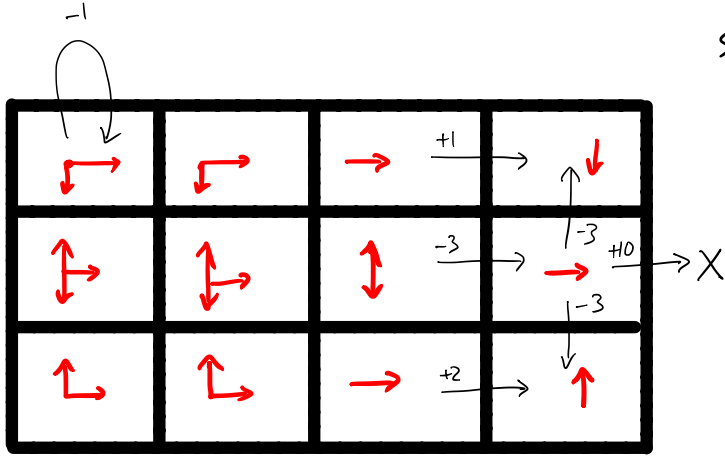
	→		
		→	
	→		

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

→	→	→	↓
↔	→	→	→
→	→	→	↑

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

		?	



Stand
Wave
Clap

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

	→		
		→	
	→		

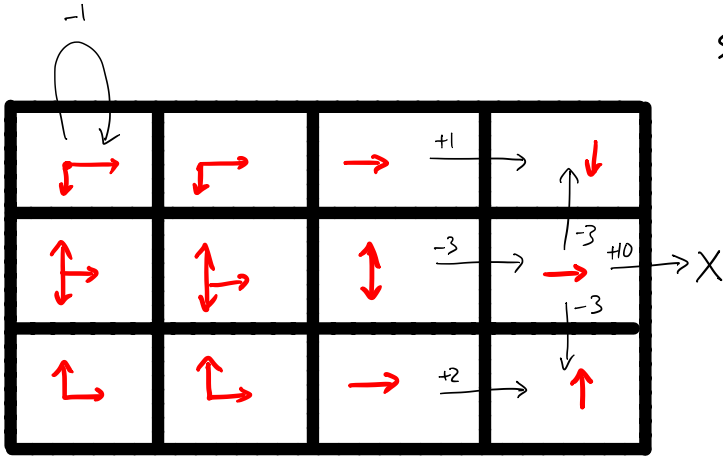
0	1	1	10
0	0	7	10
0	2	2	10

→	→	→	↓
↘	→	→	→
→	→	→	↑

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

→	→	→	↓
→	→	↓	→
→	→	→	↑

		?	
	?		



Stand
 Wave
 Clap

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

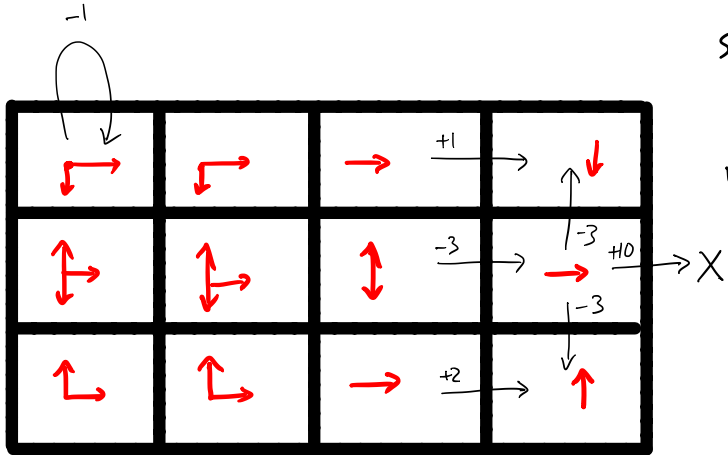
→	→	→	↓
↗	→	→	→
→	→	→	↑

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

→	→	→	↓
↗	→	→	→
→	→	→	↑

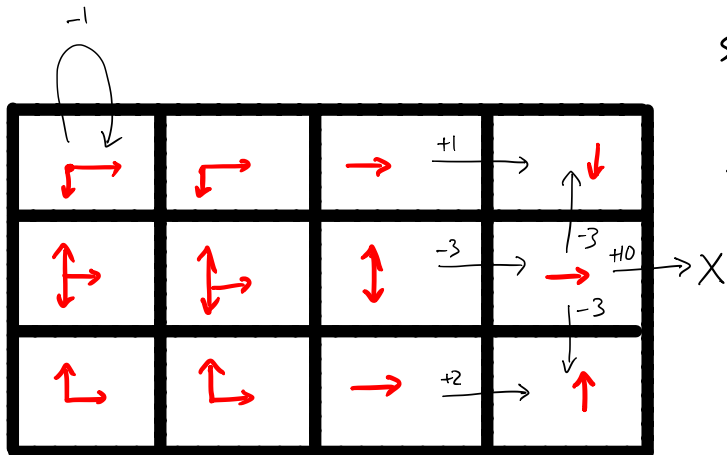
1	1	11	10
0	7	7	10
2	2	12	10

→	→	→	↓
→	→	↓	→
→	→	→	↑

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

→	→	↓	↓
→	↖	↓	→
→	→	↘	↑

= 10



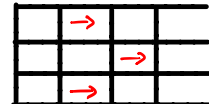
Stand
 ↓
 Clap
 ↓
 Wave

Policy Improvement

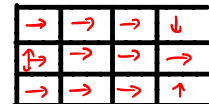
0	0	0	0
0	0	0	0
0	0	0	0



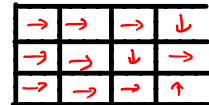
0	0	1	0
0	0	0	10
0	0	2	0



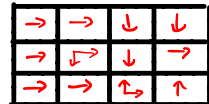
0	1	1	10
0	0	7	10
0	2	2	10



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

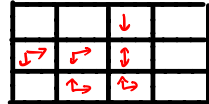


1	11	11	10
7	7	12	10
2	12	12	10



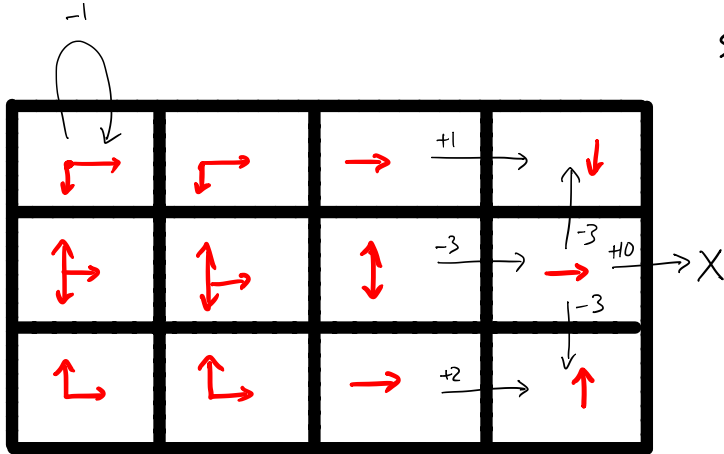
= π^*

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

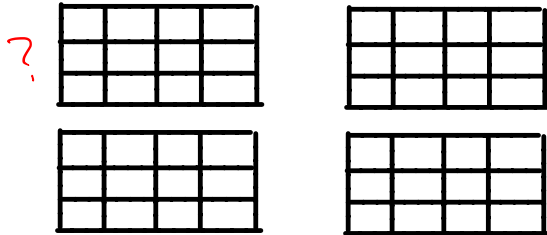


= π^*
(also)

Value Iteration



Stand
Clap
Wave



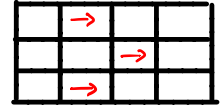
Policy Iteration

Policy Improvement

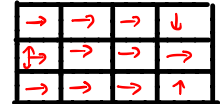
0	0	0	0
0	0	0	0
0	0	0	0



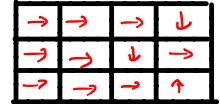
0	0	1	0
0	0	0	10
0	0	2	0



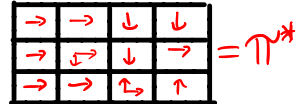
0	1	1	10
0	0	7	10
0	2	2	10



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



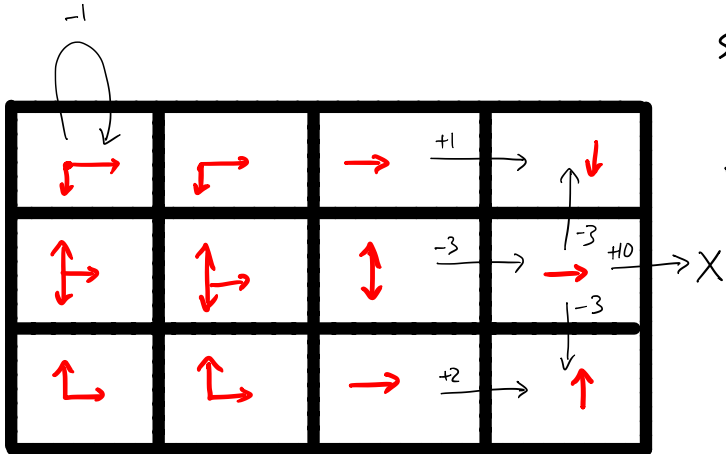
1	11	11	10
7	7	12	10
2	12	12	10



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



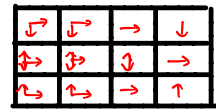
Value Iteration



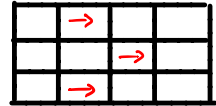
Stand
↕
Wave
Clap

Policy Improvement

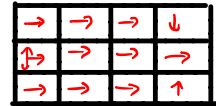
0	0	0	0
0	0	0	0
0	0	0	0



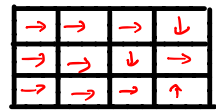
0	0	1	0
0	0	0	10
0	0	2	0



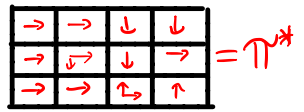
0	1	1	10
0	0	7	10
0	2	2	10



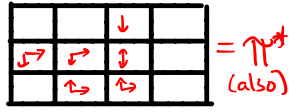
1	1	11	10
0	7	7	10
2	2	12	10



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



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



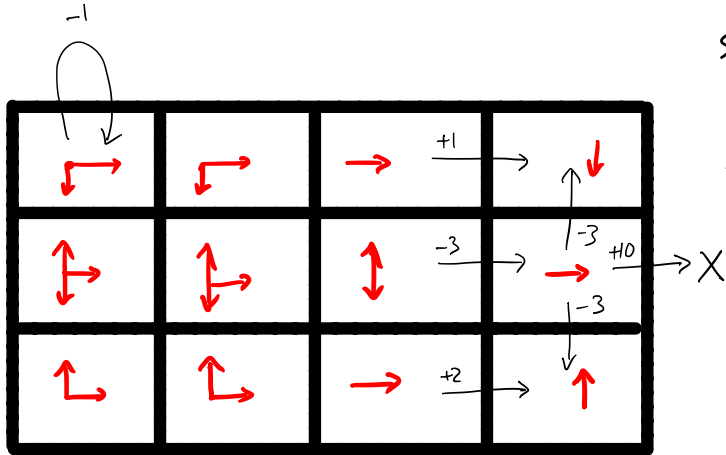
$11\frac{3}{8}$	$11\frac{1}{4}$	11	10
$11\frac{1}{2}$	$11\frac{1}{2}$	$11\frac{1}{2}$	10
$11\frac{3}{8}$	$11\frac{3}{4}$	12	10

?

?

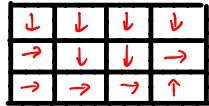
Policy Iteration

Value Iteration



Stand
↕
Wave
Clap

$11\frac{3}{8}$	$11\frac{1}{4}$	11	10
$11\frac{1}{2}$	$11\frac{1}{2}$	$11\frac{1}{2}$	10
$11\frac{3}{8}$	$11\frac{3}{4}$	12	10



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



Policy Iteration

Policy Improvement

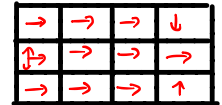
0	0	0	0
0	0	0	0
0	0	0	0



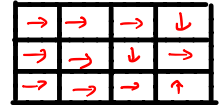
0	0	1	0
0	0	0	10
0	0	2	0



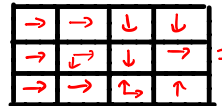
0	1	1	10
0	0	7	10
0	2	2	10



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

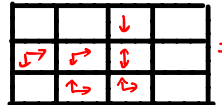


1	11	11	10
7	7	12	10
2	12	12	10



= π^*

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



= π^*
(also)

Value Iteration