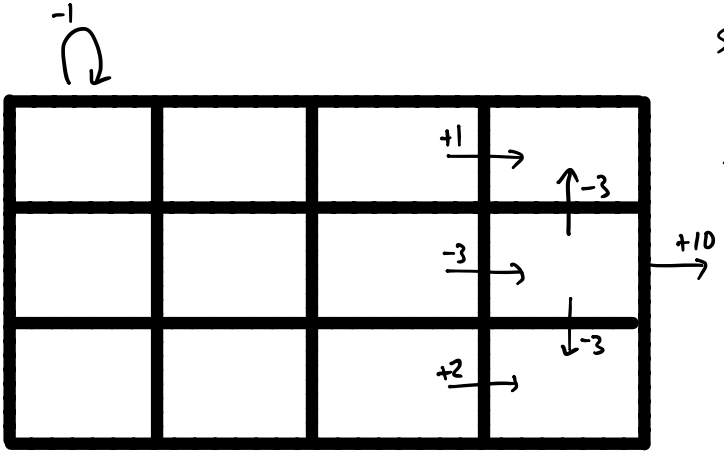


Stand

Clap

Wave

1-step

2

3

4

5

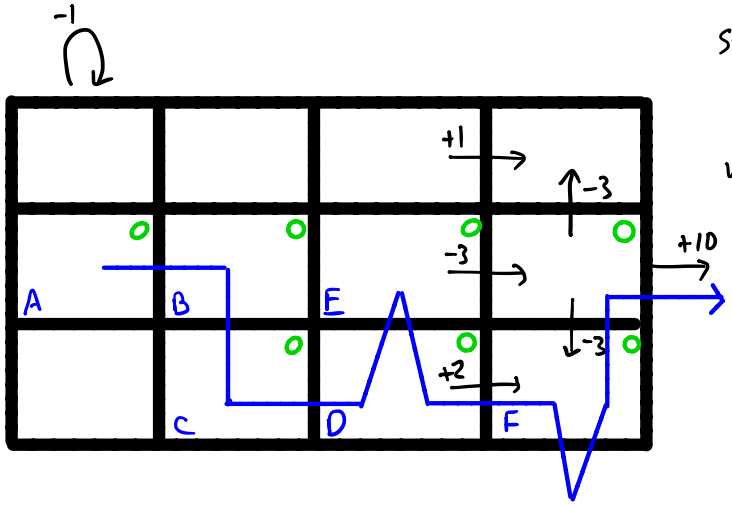
6

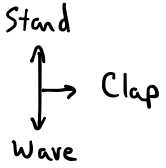
7

8

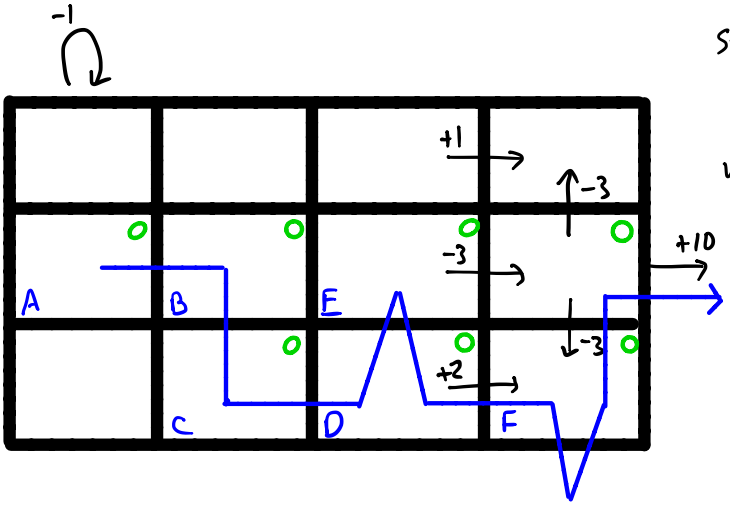
9

Review: n-step returns





	A
1-step	?
2-step	?
3	?
4	⋮
5	
6	
7	
8	
9	

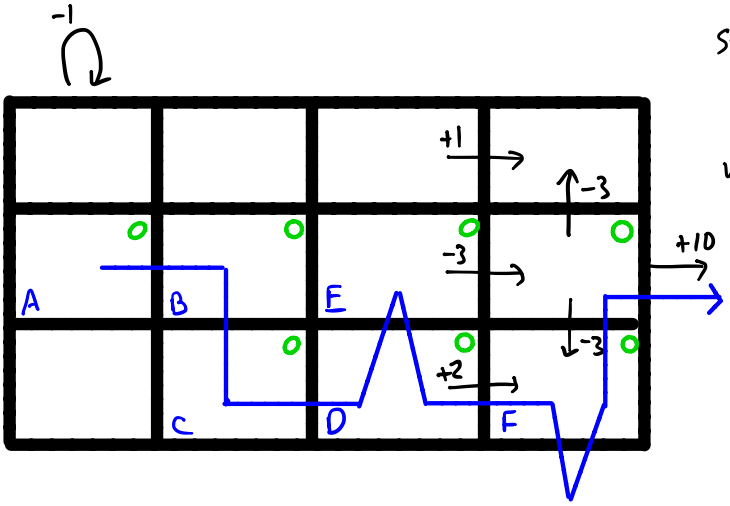


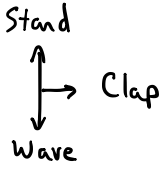
Stand

Clap

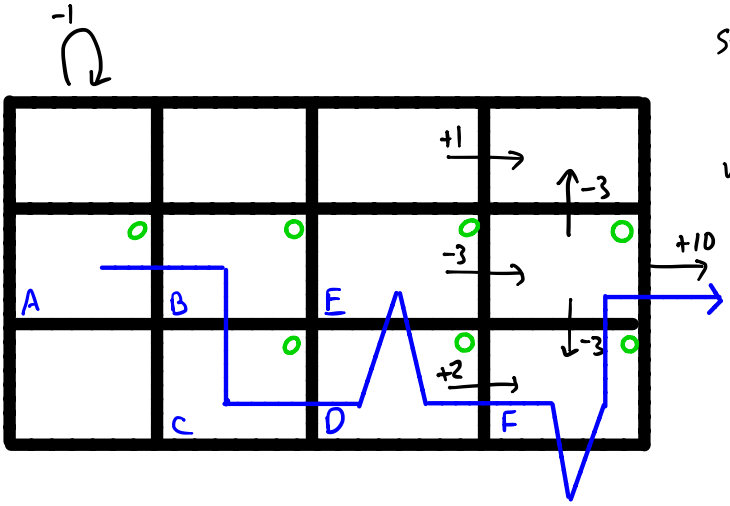
Wave

	A
1-step	0
2	0
3	0
4	0
5	0
6	28^5
7	?
8	?
9	?

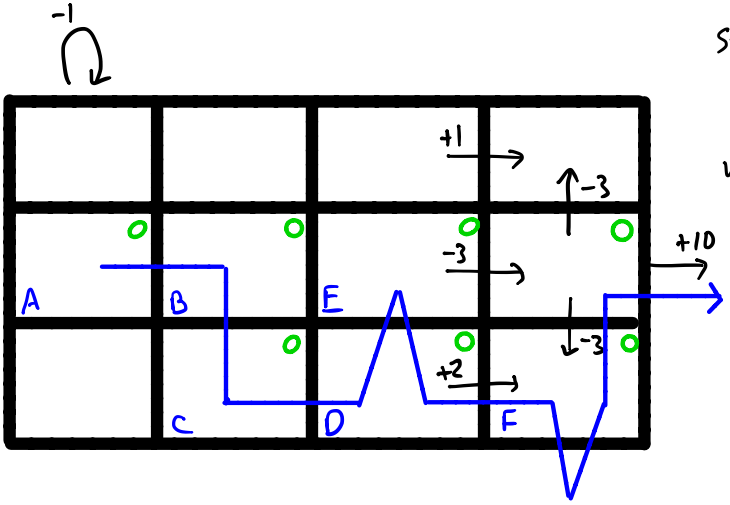




	A	B
1-step	0	?
2	0	
3	0	
4	0	
5	0	
6	$2\gamma^5$	
7	$2\gamma^5-\gamma^6$	
8	$2\gamma^5-\gamma^6$	
9	$2\gamma^5-\gamma^6+10\gamma^8$	

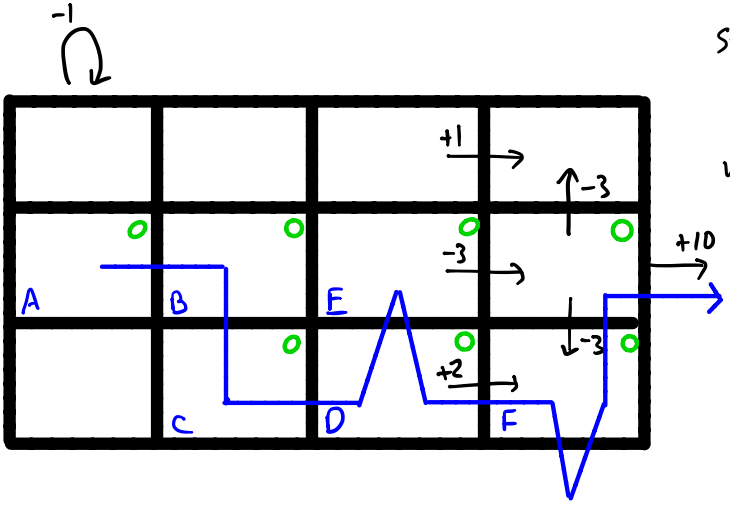


	A	B
1-step	0	0
2	0	0
3	0	0
4	0	0
5	0	$2\gamma^4$
6	$2\gamma^5$	$2\gamma^4-\gamma^5$
7	$2\gamma^5-\gamma^6$	$2\gamma^4-\gamma^5$
8	$2\gamma^5-\gamma^6$	$2\gamma^4-\gamma^5+10\gamma^7$
9	$2\gamma^5-\gamma^6+10\gamma^8$	



offline λ-return

	A	B
1-step	$0 \quad (1-\lambda)$	0
2	$0 \quad (1-\lambda)\lambda$	0
3	$0 \quad (1-\lambda)\lambda^2$	0
4	$0 \quad (1-\lambda)\lambda^3$	0
5	$0 \quad (1-\lambda)\lambda^4$	$2\gamma^4$
6	$2\gamma^5 \quad (1-\lambda)\lambda^5$	$2\gamma^4-\gamma^5$
7	$2\gamma^5-\gamma^6 \quad (1-\lambda)\lambda^6$	$2\gamma^4-\gamma^5$
8	$2\gamma^5-\gamma^6 \quad (1-\lambda)\lambda^7$	$2\gamma^4-\gamma^5+10\gamma^7$
9	$2\gamma^5-\gamma^6+10\gamma^7 \quad \lambda^8$	

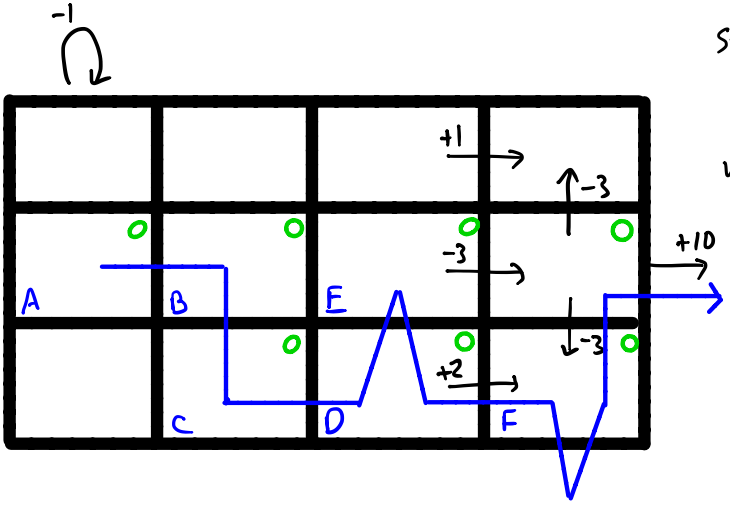


offline λ-return

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$	$2\gamma^4$	$(1-\lambda)\lambda^4$
6	$2\gamma^5$	$(1-\lambda)\lambda^5$	$2\gamma^4-\gamma^5$	$(1-\lambda)\lambda^5$
7	$2\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$2\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$2\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$2\gamma^4-\gamma^5+10\gamma^7$	λ^7
9	$2\gamma^5-\gamma^6+10\gamma^7$	λ^8		

what makes this "offline?"

Why use this bizarre weighting scheme?



Stand
Clap
Wave

offline λ-return

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$	γ^4	$(1-\lambda)\lambda^4$
6	γ^5	$(1-\lambda)\lambda^5$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^5$
7	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$\gamma^4-\gamma^5+10\gamma^7$	λ^7
9	$\gamma^5-\gamma^6+10\gamma^7$	λ^8		

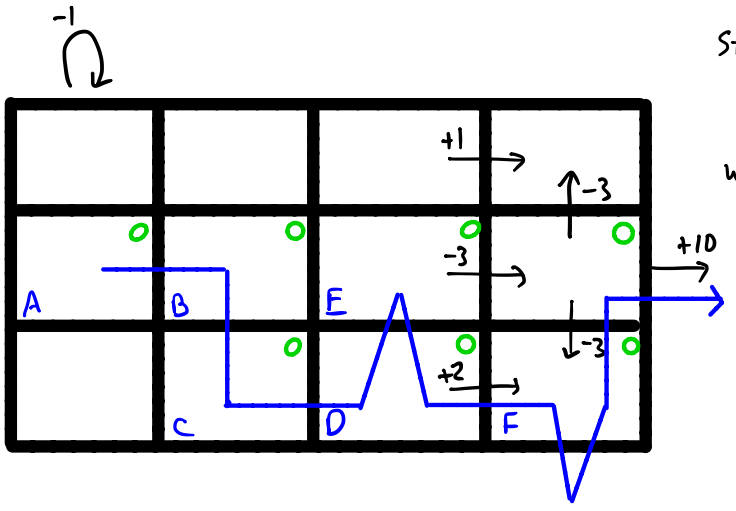
what makes this "offline?"

Why use this bizarre weighting scheme?

TD(λ) is online — why is that preferable?

TD(λ) is online — why is that preferable?

1. Updates on every step: no memory
2. Updates equally distributed in time
3. Can be applied to continuing problems: even TD(1)=MC



Stand
Clap
Wave

$w = v(A)$

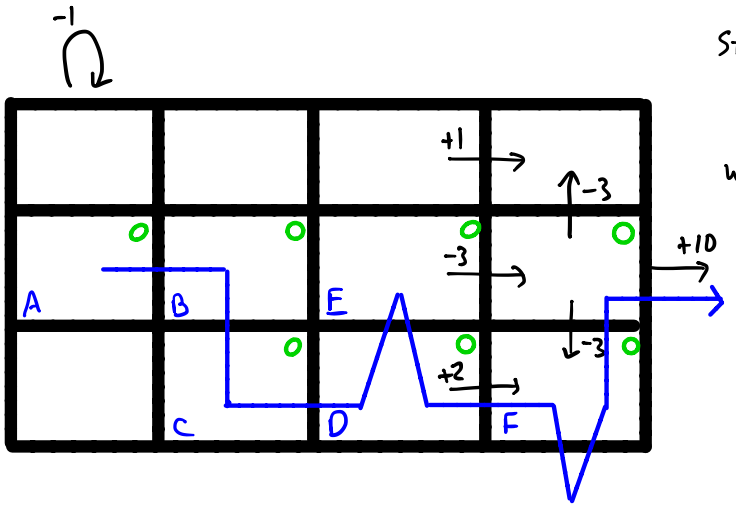
TD(λ)

$v(F)$

	z					δ
A	B	C	D	E	F	
1	0	0	0	0	0	
?						
?						
⋮						

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$ $\times\alpha\lambda$	$2\gamma^4$	$(1-\lambda)\lambda^4$
6	$2\gamma^5$	$(1-\lambda)\lambda^5$	$2\gamma^4-\gamma^5$	$(1-\lambda)\lambda^5$
7	$2\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$2\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$2\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$2\gamma^4-\gamma^5+10\gamma^7$	λ^7
9	$2\gamma^5-\gamma^6+10\gamma^8$	λ^8		

$$\delta = R + \gamma v(s') - v(s)$$
$$w = w + \alpha \delta z$$



Stand
Clap
Wave

TD(λ)

$w = v(A)$ $v(F)$

z

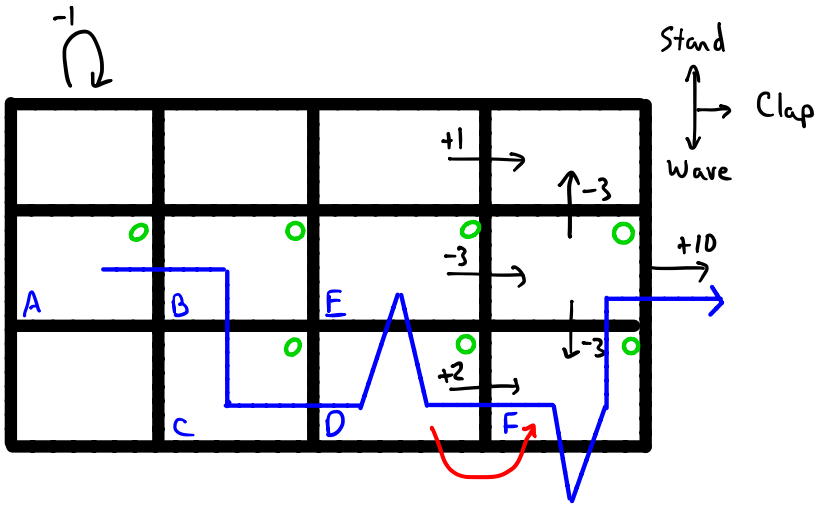
A	B	C	D	E	F
1	0	0	0	0	0
$\gamma\lambda$					
$(\gamma\lambda)^2$					
⋮					
$(\gamma\lambda)^4$	?	?	---		

δ

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$	γ^4	$(1-\lambda)\lambda^4$
6	γ^5	$(1-\lambda)\lambda^5$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^5$
7	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$\gamma^4-\gamma^5+10\gamma^7$	λ^7
9	$\gamma^5-\gamma^6+10\gamma^8$	λ^8		

$\times\gamma\lambda$

$$\delta = R + \gamma v(s') - v(s)$$
$$w = w + \alpha\delta z$$



z

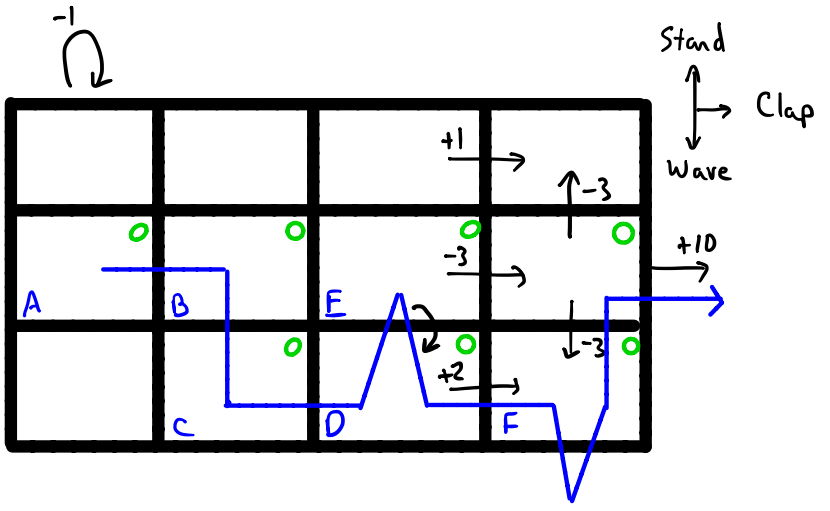
A	B	C	D	E	F	δ
1	0	0	0	0	0	?
$\gamma\lambda$						
$(\gamma\lambda)^2$						
⋮						⋮
$(\gamma\lambda)^4$	$(\gamma\lambda)^3$	$(\gamma\lambda)^2$	$(\gamma\lambda)$	1	0	?
?						

TD(λ)

$w = v(A)$ $v(F)$

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$ $\times\gamma\lambda$	γ^4	$(1-\lambda)\lambda^4$
6	γ^5	$(1-\lambda)\lambda^5$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^5$
7	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$\gamma^4-\gamma^5+10\gamma^7$	λ^7
9	$\gamma^5-\gamma^6+10\gamma^7$	λ^8		

$$\delta = R + \gamma v(s') - v(s)$$
$$w = w + \alpha\delta z$$



TD(λ)

$w = v(A)$?

$v(F)$

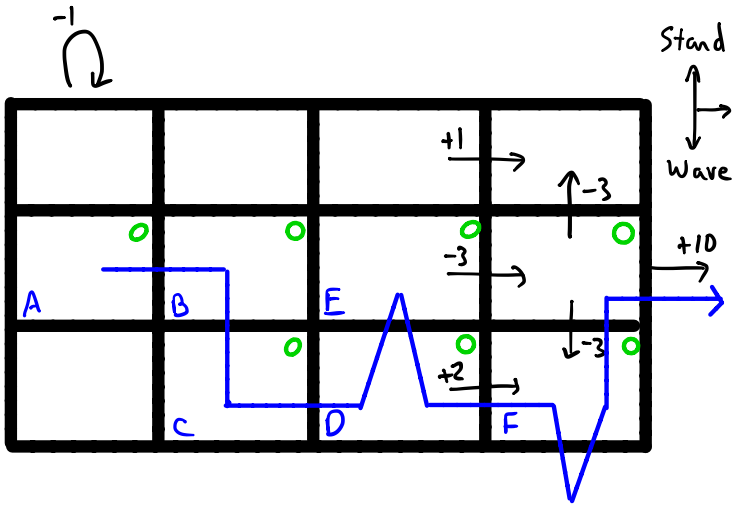
	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$ $\times\gamma\lambda$	γ^4	$(1-\lambda)\lambda^4$
6	γ^5	$(1-\lambda)\lambda^5$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^5$
7	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$\gamma^4-\gamma^5+10\gamma^7$	λ^7
9	$\gamma^5-\gamma^6+10\gamma^8$	λ^8		

z

A	B	C	D	E	F	δ
1	0	0	0	0	0	0
$\gamma\lambda$						0
$(\gamma\lambda)^2$						;
⋮						
$(\gamma\lambda)^4$	$(\gamma\lambda)^3$	$(\gamma\lambda)^2$	$(\gamma\lambda)$	1	0	0
$(\gamma\lambda)^5$	$(\gamma\lambda)^4$	$(\gamma\lambda)^3$	$(\gamma\lambda)^2+1$	$\gamma\lambda$	0	2

↑ accumulating trace

$$\delta = R + \gamma v(s') - v(s)$$
$$w = w + \alpha\delta z$$



TD(λ)

$w = v(A)$ $\quad 2\alpha(\delta z)^5$ $\quad v(B)$ $\quad 2\alpha(\delta z)^4$ $\quad v(F)$ $\quad 0$

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$ $\times\gamma\lambda$	$2\gamma^4$	$(1-\lambda)\lambda^4$
6	$2\gamma^5$	$(1-\lambda)\lambda^5$	$2\gamma^4-\gamma^5$	$(1-\lambda)\lambda^5$
7	$2\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$2\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$2\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$2\gamma^4-\gamma^5+10\gamma^7$	λ^7
9	$2\gamma^5-\gamma^6+10\gamma^8$	λ^8		

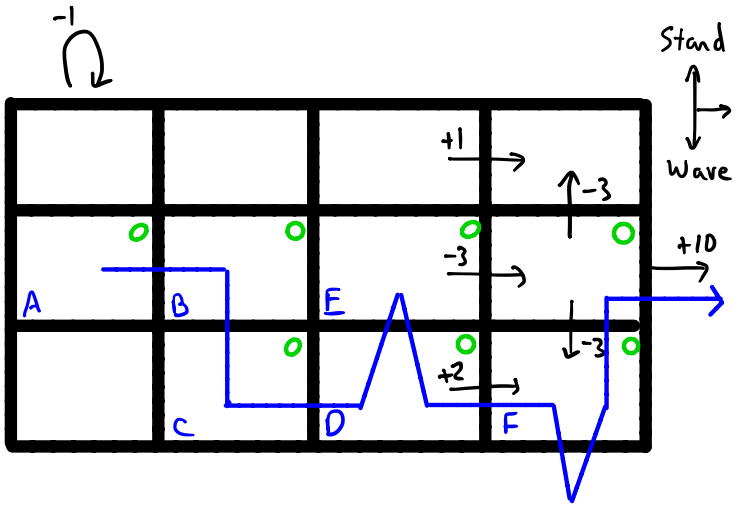
z

A	B	C	D	E	F
1	0	0	0	0	0
$\gamma\lambda$					
$(\delta\lambda)^2$		<snip>			
$(\delta\lambda)^5$	$(\delta\lambda)^4$	$(\delta\lambda)^3$	$(\delta\lambda)^2+1$	$\delta\lambda$	0
?	----				

δ

0
⋮
2

$$\delta = R + \gamma v(s') - v(s)$$
$$w = w + \alpha\delta z$$



TD(λ)

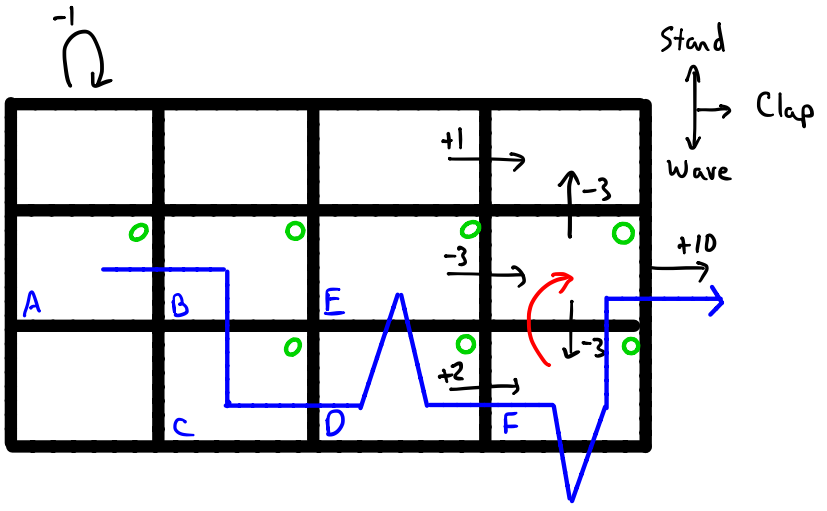
	w=v(A)	v(B)	v(F)
	$2\alpha(\delta\lambda)^5$	$2\alpha(\delta\lambda)^4$	0
	?		?

z

A	B	C	D	E	F	δ
1	0	0	0	0	0	0
$\gamma\lambda$						⋮
$(\delta\lambda)^2$		<snip>				
⋮						
$(\delta\lambda)^5$	$(\delta\lambda)^4$	$(\delta\lambda)^3$	$(\delta\lambda)^2+1$	$\delta\lambda$	0	2
$(\delta\lambda)^6$	$(\delta\lambda)^5$	$(\delta\lambda)^4$	$(\delta\lambda)^3+\delta\lambda$	$(\delta\lambda)^2$	1	-1

	A	B
1-step	0 $(1-\lambda)$	0 $(1-\lambda)$
2	0 $(1-\lambda)\lambda$	0 $(1-\lambda)\lambda$
3	0 $(1-\lambda)\lambda^2$	0 $(1-\lambda)\lambda^2$
4	0 $(1-\lambda)\lambda^3$	0 $(1-\lambda)\lambda^3$
5	0 $(1-\lambda)\lambda^4$ $\times\alpha\lambda$	$2\gamma^4$ $(1-\lambda)\lambda^4$
6	$2\gamma^5$ $(1-\lambda)\lambda^5$	$2\gamma^4-\gamma^5$ $(1-\lambda)\lambda^5$
7	$2\gamma^5-\gamma^6$ $(1-\lambda)\lambda^6$	$2\gamma^4-\gamma^5$ $(1-\lambda)\lambda^6$
8	$2\gamma^5-\gamma^6$ $(1-\lambda)\lambda^7$	$2\gamma^4-\gamma^5+10\gamma^7$ λ^7
9	$2\gamma^5-\gamma^6+10\gamma^8$ λ^8	

$$\delta = R + \gamma v(s') - v(s)$$
$$w = w + \alpha\delta z$$



TD(λ)

$w = v(A)$

$2\alpha(\delta\lambda)^5$

$2\alpha(\delta\lambda)^5 - \alpha(\delta\lambda)^6$

?

	$v(F)$
	0
	$-\alpha$

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$ $\times\delta\lambda$	δ^4	$(1-\lambda)\lambda^4$
6	δ^5	$(1-\lambda)\lambda^5$	$\delta^4-\delta^5$	$(1-\lambda)\lambda^5$
7	$\delta^5-\delta^6$	$(1-\lambda)\lambda^6$	$\delta^4-\delta^5$	$(1-\lambda)\lambda^6$
8	$\delta^5-\delta^6$	$(1-\lambda)\lambda^7$	$\delta^4-\delta^5+10\delta^7$	λ^7
9	$\delta^5-\delta^6+10\delta^8$	λ^8		

z

A	B	C	D	E	F
1	0	0	0	0	0
$\delta\lambda$					
$(\delta\lambda)^2$					
⋮		<snip>			
$(\delta\lambda)^5$	$(\delta\lambda)^4$	$(\delta\lambda)^3$	$(\delta\lambda)^2+1$	$\delta\lambda$	0
$(\delta\lambda)^6$	$(\delta\lambda)^5$	$(\delta\lambda)^4$	$(\delta\lambda)^3+\delta\lambda$	$(\delta\lambda)^2$	1

δ

0

⋮

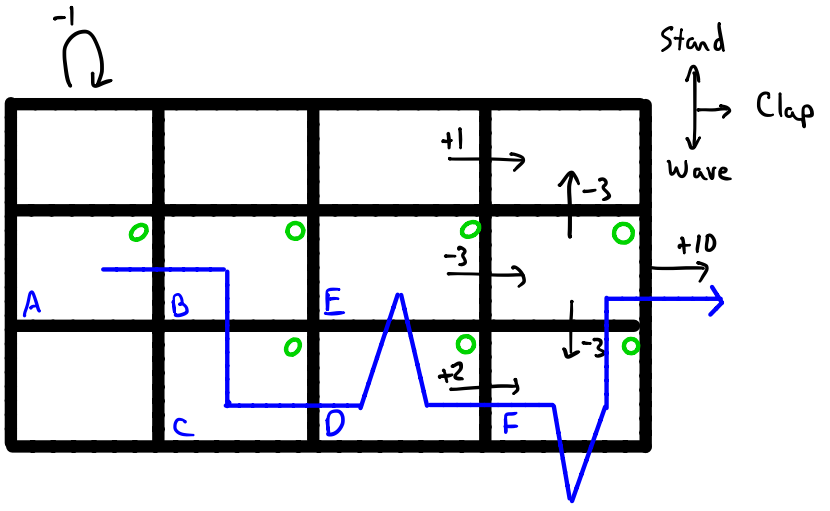
2

-1

?

$$\delta = R + \gamma v(s') - v(s)$$

$$w = w + \alpha\delta z$$



TD(λ)

$w = v(A)$

$2\alpha(\delta\lambda)^5$

$2\alpha(\delta\lambda)^5 - \alpha(\delta\lambda)^6$

$2\alpha(\delta\lambda)^5 - \alpha(\delta\lambda)^6 + \alpha^2(\delta\lambda)^7$

$v(F)$

0

$-\alpha$

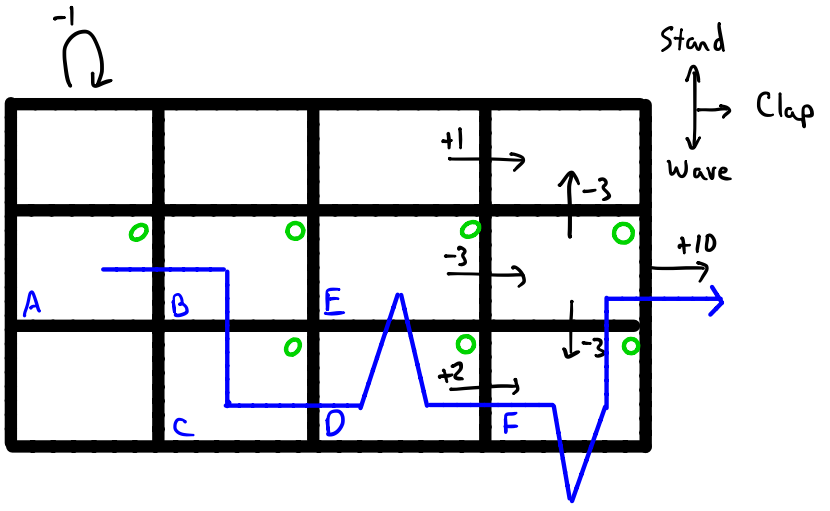
	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$ $\times\delta\lambda$	δ^4	$(1-\lambda)\lambda^4$
6	δ^5	$(1-\lambda)\lambda^5$	$\delta^4-\delta^5$	$(1-\lambda)\lambda^5$
7	$\delta^5-\delta^6$	$(1-\lambda)\lambda^6$	$\delta^4-\delta^5$	$(1-\lambda)\lambda^6$
8	$\delta^5-\delta^6$	$(1-\lambda)\lambda^7$	$\delta^4-\delta^5+10\delta^7$	λ^7
9	$\delta^5-\delta^6+10\delta^8$	λ^8		

A	B	C	D	E	F	δ
z						
1	0	0	0	0	0	0
$\delta\lambda$						⋮
$(\delta\lambda)^2$						
⋮		<snip>				
$(\delta\lambda)^5$	$(\delta\lambda)^4$	$(\delta\lambda)^3$	$(\delta\lambda)^2+1$	$\delta\lambda$	0	2
$(\delta\lambda)^6$	$(\delta\lambda)^5$	$(\delta\lambda)^4$	$(\delta\lambda)^3+\delta\lambda$	$(\delta\lambda)^2$	1	-1
$(\delta\lambda)^7$						α

different from off line case

$$\delta = R + \gamma v(s') - v(s)$$

$$w = w + \alpha\delta z$$



TD(λ)

$w = v(A)$

$2\alpha(\delta\lambda)^5$

$2\alpha(\delta\lambda)^5 - \alpha(\delta\lambda)^6$

$2\alpha(\delta\lambda)^5 - \alpha(\delta\lambda)^6 + \alpha^2(\delta\lambda)^7$

⋮

Builds up incrementally to almost same result as offline forward view

$v(F)$

0

$-\alpha$

TD(0) ↔ TD(1)

TD ··· λ ··· MC

	A		B	
1-step	0	$(1-\lambda)$	0	$(1-\lambda)$
2	0	$(1-\lambda)\lambda$	0	$(1-\lambda)\lambda$
3	0	$(1-\lambda)\lambda^2$	0	$(1-\lambda)\lambda^2$
4	0	$(1-\lambda)\lambda^3$	0	$(1-\lambda)\lambda^3$
5	0	$(1-\lambda)\lambda^4$	$\gamma\lambda$	$(1-\lambda)\lambda^4$
6	γ^5	$(1-\lambda)\lambda^5$	γ^4	$(1-\lambda)\lambda^5$
7	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^6$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
8	$\gamma^5-\gamma^6$	$(1-\lambda)\lambda^7$	$\gamma^4-\gamma^5$	$(1-\lambda)\lambda^6$
9	$\gamma^5-\gamma^6+10\gamma^7$	λ^8	$\gamma^4-\gamma^5+10\gamma^7$	λ^7

	z					δ
A	B	C	D	E	F	
1	0	0	0	0	0	0
$\gamma\lambda$						⋮
$(\gamma\lambda)^2$						
⋮		<snip>				
$(\gamma\lambda)^5$	$(\gamma\lambda)^4$	$(\gamma\lambda)^3$	$(\gamma\lambda)^2+1$	$\gamma\lambda$	0	2
$(\gamma\lambda)^6$	$(\gamma\lambda)^5$	$(\gamma\lambda)^4$	$(\gamma\lambda)^3+\gamma\lambda$	$(\gamma\lambda)^2$	1	-1
$(\gamma\lambda)^7$						α

different from offline case

$$\delta = R + \gamma v(s') - v(s)$$

$$w = w + \alpha\delta z$$

offline λ-return

TD(λ)

offline λ-return

TD(λ)

(Truncated λ-return
Online λ-return)

True Online TD(λ)

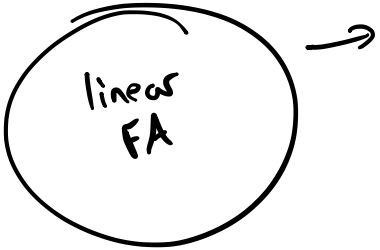
offline λ-return

TD(λ)

Truncated λ-return

Online λ-return

True Online TD(λ)



Online λ-return

$$h = 1: \quad \mathbf{w}_1^1 \doteq \mathbf{w}_0^1 + \alpha \left[G_{0:1}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^1) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^1), \quad \longleftarrow \ ?$$

$$\begin{aligned} h = 2: \quad \mathbf{w}_1^2 &\doteq \mathbf{w}_0^2 + \alpha \left[G_{0:2}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^2) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^2), \\ \mathbf{w}_2^2 &\doteq \mathbf{w}_1^2 + \alpha \left[G_{1:2}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^2) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^2), \end{aligned}$$

$$\begin{aligned} h = 3: \quad \mathbf{w}_1^3 &\doteq \mathbf{w}_0^3 + \alpha \left[G_{0:3}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^3) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^3), \\ \mathbf{w}_2^3 &\doteq \mathbf{w}_1^3 + \alpha \left[G_{1:3}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^3) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^3), \\ \mathbf{w}_3^3 &\doteq \mathbf{w}_2^3 + \alpha \left[G_{2:3}^{\lambda} - \hat{v}(S_2, \mathbf{w}_2^3) \right] \nabla \hat{v}(S_2, \mathbf{w}_2^3). \end{aligned}$$

How differs from TD(λ)?

Online λ-return

$$h = 1: \quad \mathbf{w}_1^1 \doteq \mathbf{w}_0^1 + \alpha \left[G_{0:1}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^1) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^1),$$

← No change

$$h = 2: \quad \mathbf{w}_1^2 \doteq \mathbf{w}_0^2 + \alpha \left[G_{0:2}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^2) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^2),$$
$$\mathbf{w}_2^2 \doteq \mathbf{w}_1^2 + \alpha \left[G_{1:2}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^2) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^2),$$

$$h = 3: \quad \mathbf{w}_1^3 \doteq \mathbf{w}_0^3 + \alpha \left[G_{0:3}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^3) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^3),$$
$$\mathbf{w}_2^3 \doteq \mathbf{w}_1^3 + \alpha \left[G_{1:3}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^3) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^3),$$
$$\mathbf{w}_3^3 \doteq \mathbf{w}_2^3 + \alpha \left[G_{2:3}^{\lambda} - \hat{v}(S_2, \mathbf{w}_2^3) \right] \nabla \hat{v}(S_2, \mathbf{w}_2^3).$$

How differs from TD(λ)?

Online λ-return

$$h = 1: \quad \mathbf{w}_1^1 \doteq \mathbf{w}_0^1 + \alpha \left[G_{0:1}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^1) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^1),$$

← No change

$$h = 2: \quad \cancel{\mathbf{w}_1^2 \doteq \mathbf{w}_0^2 + \alpha \left[G_{0:2}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^2) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^2),}$$

$$\mathbf{w}_2^2 \doteq \cancel{\mathbf{w}_1^2} + \alpha \left[G_{\cancel{1:2}}^{\lambda} - \hat{v}(S_1, \cancel{\mathbf{w}_1^2}) \right] \nabla \hat{v}(S_1, \cancel{\mathbf{w}_1^2}),$$

$\mathbf{w}_1'$ $\quad 0{:}2 \quad$ $\mathbf{w}_1'$ $\quad \mathbf{w}_1'$

$G_{0:2}^{\lambda}$ with weights $(1-\lambda)$, $(1-\lambda)\lambda$

Online truncates: $(1-\lambda)$, λ

$$h = 3: \quad \mathbf{w}_1^3 \doteq \mathbf{w}_0^3 + \alpha \left[G_{0:3}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^3) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^3),$$

$$\mathbf{w}_2^3 \doteq \mathbf{w}_1^3 + \alpha \left[G_{1:3}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^3) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^3),$$

$$\mathbf{w}_3^3 \doteq \mathbf{w}_2^3 + \alpha \left[G_{2:3}^{\lambda} - \hat{v}(S_2, \mathbf{w}_2^3) \right] \nabla \hat{v}(S_2, \mathbf{w}_2^3).$$

How differs from TD(λ)?

Online λ-return

$$h = 1: \quad \mathbf{w}_1^1 \doteq \mathbf{w}_0^1 + \alpha \left[G_{0:1}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^1) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^1),$$

← No change

$$h = 2: \quad \cancel{\mathbf{w}_1^2 \doteq \mathbf{w}_0^2 + \alpha \left[G_{0:2}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^2) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^2),}$$

$$\mathbf{w}_2^2 \doteq \cancel{\mathbf{w}_1^2} + \alpha \left[\cancel{G_{1:2}^{\lambda}} - \hat{v}(S_1, \cancel{\mathbf{w}_1^2}) \right] \nabla \hat{v}(S_1, \cancel{\mathbf{w}_1^2}),$$

$\mathbf{w}_1^1$ $\quad$ $0{:}2$ $\quad$ $\mathbf{w}_1^1$ $\quad$ $\mathbf{w}_1^1$

$G_{0:2}^{\lambda}$ with weights $(1-\lambda)$, $(1-\lambda)\lambda$

(online truncates: $(1-\lambda)$, λ)

$$h = 3: \quad \cancel{\mathbf{w}_1^3 \doteq \mathbf{w}_0^3 + \alpha \left[G_{0:3}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^3) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^3),}$$

$$\cancel{\mathbf{w}_2^3 \doteq \mathbf{w}_1^3 + \alpha \left[G_{1:3}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^3) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^3),}$$

$$\mathbf{w}_3^3 \doteq \cancel{\mathbf{w}_2^3} + \alpha \left[\cancel{G_{2:3}^{\lambda}} - \hat{v}(S_2, \cancel{\mathbf{w}_2^3}) \right] \nabla \hat{v}(S_2, \cancel{\mathbf{w}_2^3}).$$

$\mathbf{w}_2^2$ $\quad$ $0{:}3$ $\quad$ $\mathbf{w}_2^2$ $\quad$ $\mathbf{w}_2^2$

$G_{0:3}^{\lambda}$ with weights $(1-\lambda)$, $(1-\lambda)\lambda$, $(1-\lambda)\lambda^2$

How differs from TD(λ)?

Online λ-return

$$h=1: \quad \mathbf{w}_1^1 \doteq \mathbf{w}_0^1 + \alpha \left[G_{0:1}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^1) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^1),$$

$$\begin{aligned} h=2: \quad \mathbf{w}_1^2 &\doteq \mathbf{w}_0^2 + \alpha \left[G_{0:2}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^2) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^2), \\ \mathbf{w}_2^2 &\doteq \mathbf{w}_1^2 + \alpha \left[G_{1:2}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^2) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^2), \end{aligned}$$

$$\begin{aligned} h=3: \quad \mathbf{w}_1^3 &\doteq \mathbf{w}_0^3 + \alpha \left[G_{0:3}^{\lambda} - \hat{v}(S_0, \mathbf{w}_0^3) \right] \nabla \hat{v}(S_0, \mathbf{w}_0^3), \\ \mathbf{w}_2^3 &\doteq \mathbf{w}_1^3 + \alpha \left[G_{1:3}^{\lambda} - \hat{v}(S_1, \mathbf{w}_1^3) \right] \nabla \hat{v}(S_1, \mathbf{w}_1^3), \\ \mathbf{w}_3^3 &\doteq \mathbf{w}_2^3 + \alpha \left[G_{2:3}^{\lambda} - \hat{v}(S_2, \mathbf{w}_2^3) \right] \nabla \hat{v}(S_2, \mathbf{w}_2^3). \end{aligned}$$

True Online TD(λ) computes this fully incrementally using the "Dutch" trace (see Sutton's slides for illustration)

offline λ-return

TD(λ)

Truncated λ-return

Online λ-return

True Online TD(λ)