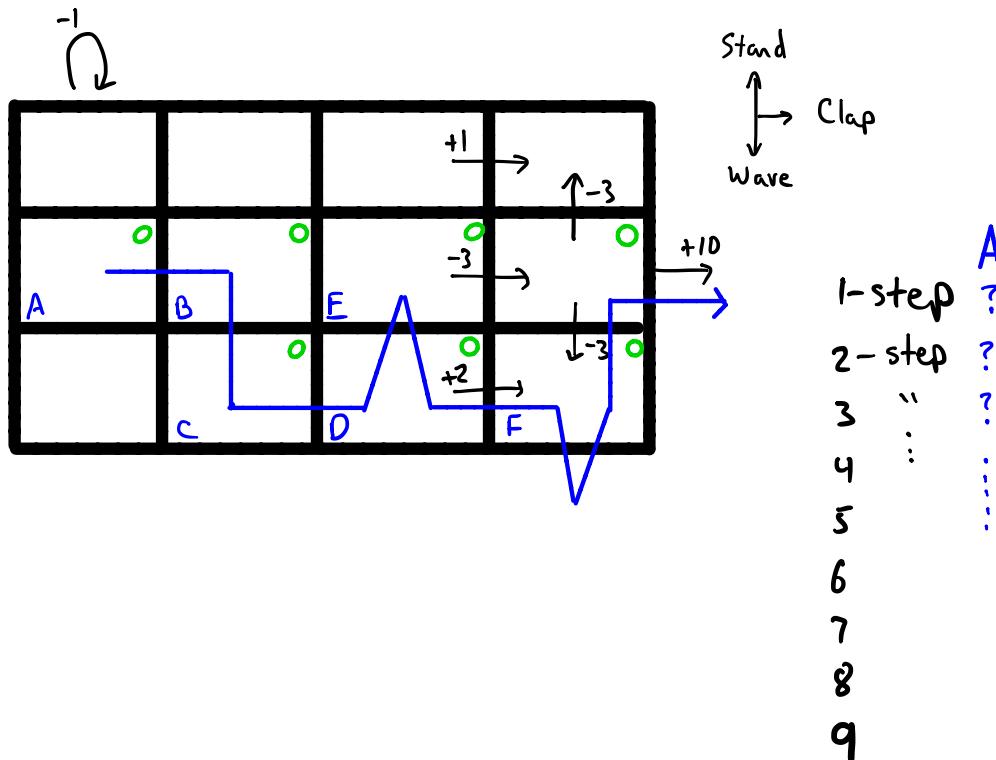


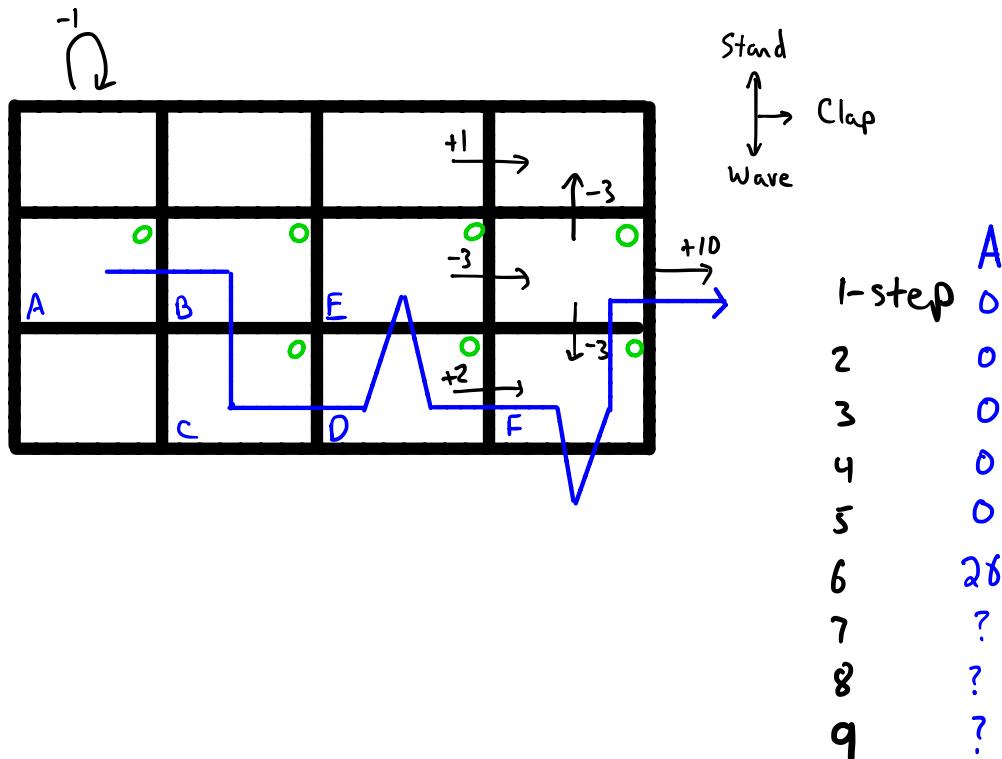
Stand
Clap
Wave

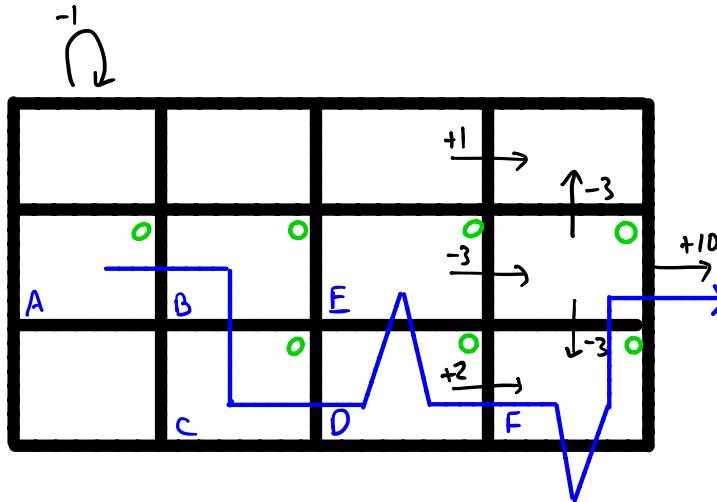
1-step

2
3
4
5
6
7
8
9

Review: n-step returns





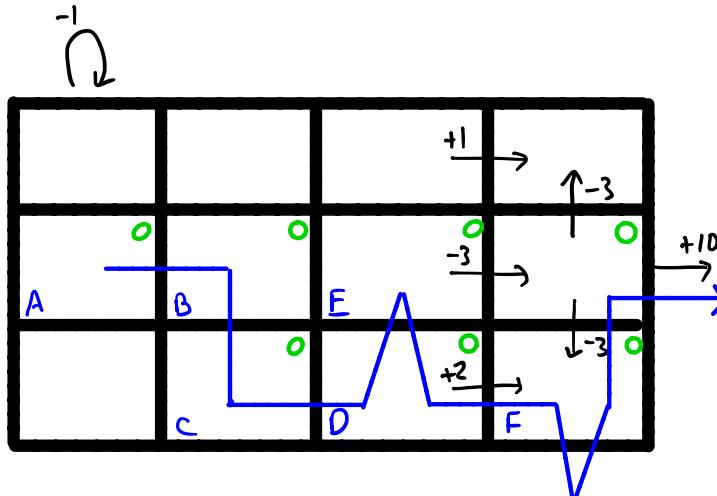


Stand
Wave

Clap

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

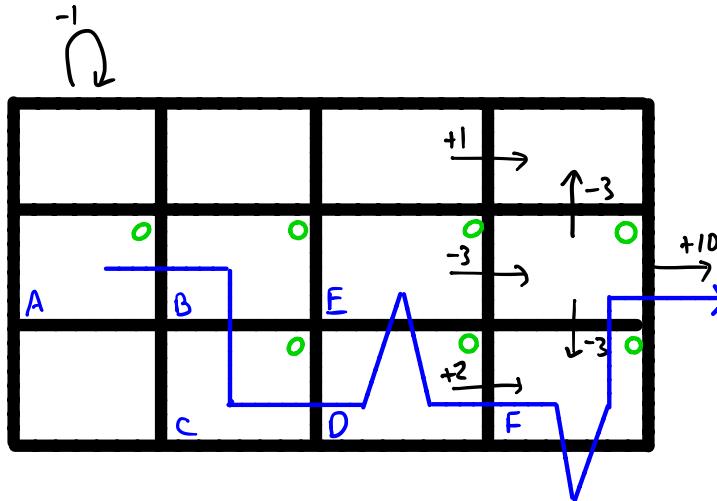
B
?



Stand
Clap
Wave

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

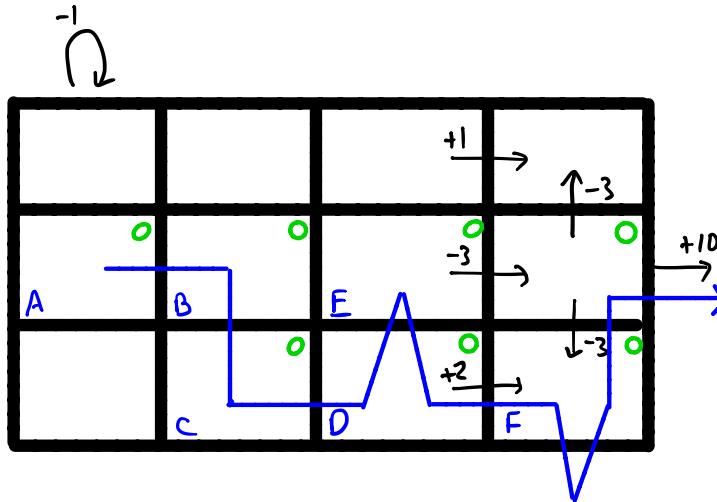
A
 B
 0
 0
 0
 0
 $2\delta^4$
 $2\delta^1 - \delta^5$
 $2\delta^4 - \delta^5$
 $2\delta^4 - \delta^5 + 10\delta^7$



Stand
Clap
Wave

offline λ -return

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



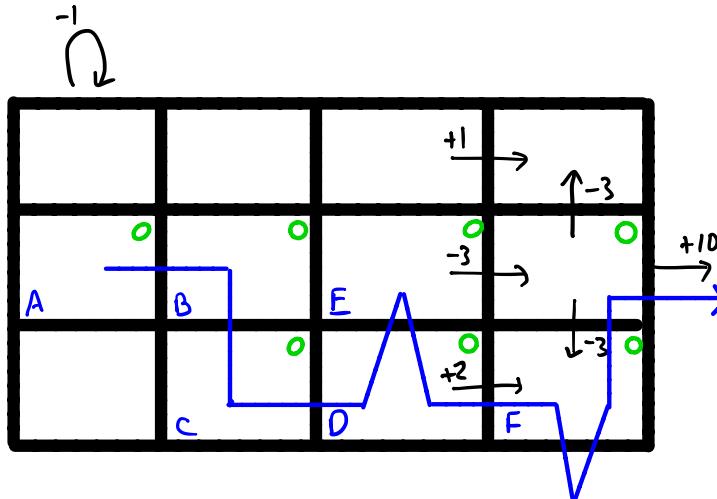
Stand
 Clap
 Wave

offline λ -return

what makes this "offline?"

Why use this bizarre weighting scheme?

	A	B
1-step	$(1-\lambda)$	$(1-\lambda)$
2	$(1-\lambda)\lambda$	$(1-\lambda)\lambda$
3	$(1-\lambda)^2\lambda$	$(1-\lambda)^2\lambda$
4	$(1-\lambda)^3\lambda$	$(1-\lambda)^3\lambda$
5	$(1-\lambda)^4\lambda$	$(1-\lambda)^4\lambda$
6	$(1-\lambda)^5\lambda$	$(1-\lambda)^5\lambda$
7	$(1-\lambda)^6\lambda$	$(1-\lambda)^6\lambda$
8	$(1-\lambda)^7\lambda$	$(1-\lambda)^7\lambda$
9	$(1-\lambda)^8\lambda$	$(1-\lambda)^8\lambda$



Stand
Clap
Wave

offline λ -return

what makes this "offline?"

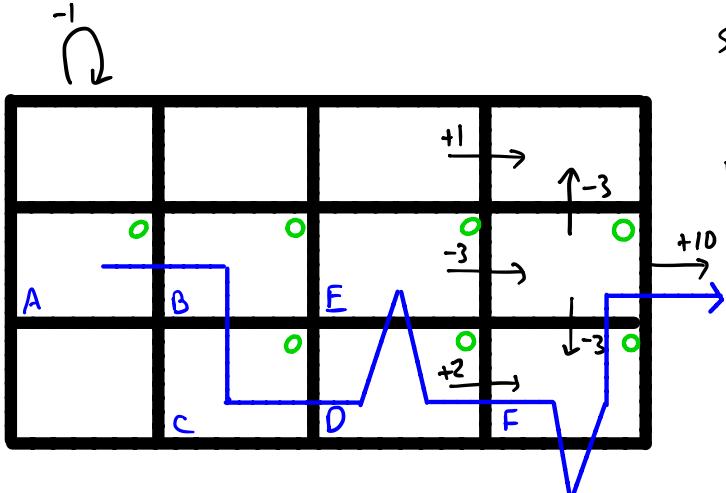
Why use this bizarre weighting scheme?

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

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

$TD(\lambda)$ 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 $TD(1) = MC$



z

A	B	C	D	E	F
0	0	0	0	0	0
?	?	:			

Stand
Clap
Wave

$$w = v(A)$$

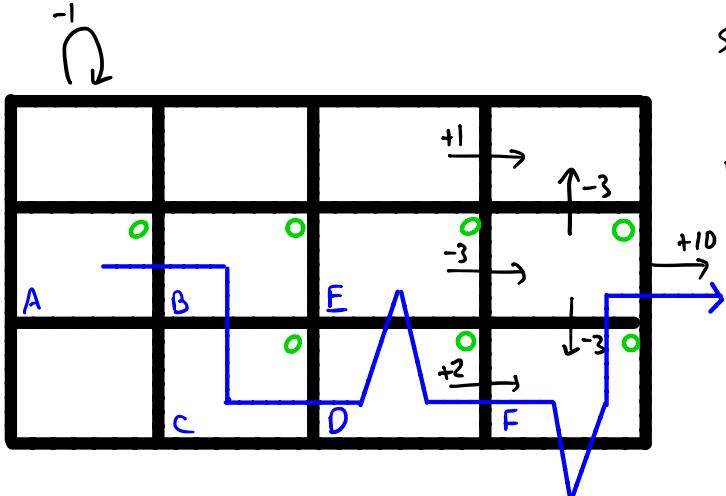
$$TD(\lambda)$$

$$v(F)$$

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

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

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



A	B	C	D	E	F
1	0	0	0	0	0
δ^1	$(\delta^1)^2$?	?	---	
$(\delta^1)^3$	$(\delta^1)^4$				

Stand \uparrow
 Clap \downarrow
 Wave

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

TD(z)

$$w$$

$$v(A)$$

$$v(F)$$

$$\delta$$

$$z$$

$$A$$

$$B$$

$$C$$

$$D$$

$$E$$

$$F$$

$$\delta$$

$$z$$

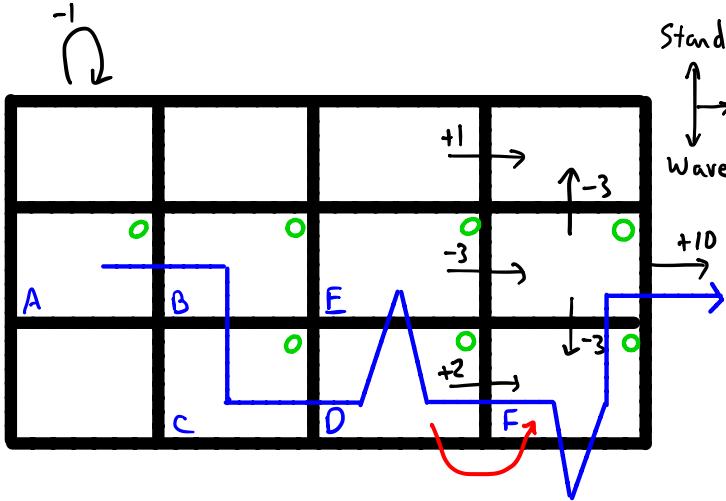
$$A$$

$$B$$

$$C$$

$$D$$

$$E$$



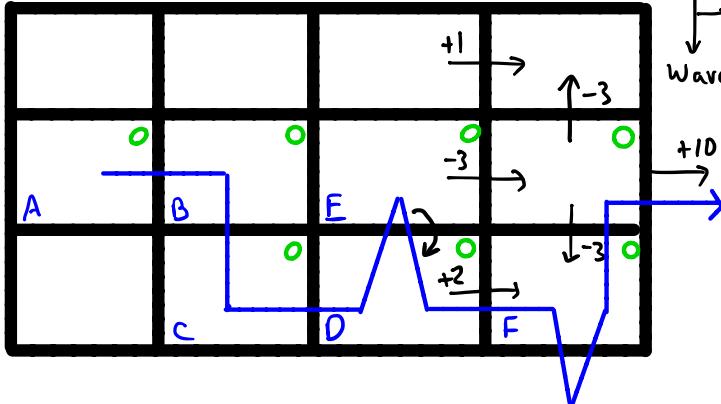
$$\begin{array}{ccccccc}
 & & z & & & & \delta \\
 A & B & C & D & E & F & ? \\
 | & 0 & 0 & 0 & 0 & 0 & ? \\
 \gamma_1 & & & & & & \vdots \\
 (\gamma_1)^2 & & & & & & \\
 \vdots & & & & & & \\
 (\gamma_1)^4 & (\gamma_1)^3 & (\gamma_1)^2 & (\gamma_1) & 1 & 0 & ? \\
 \end{array}$$

$$\begin{array}{l}
 TD(\lambda) \\
 w = v(A) \\
 v(F)
 \end{array}$$

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

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

-1



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

↑
accumulating
trace

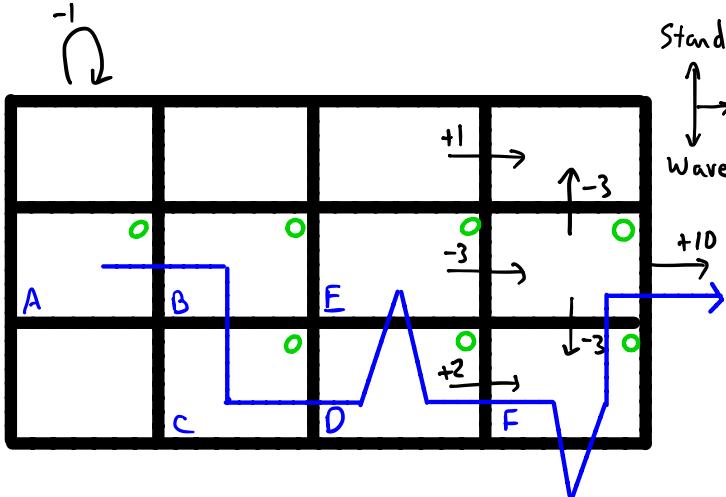
$$w = v(A) \\ ?$$

$TD(\lambda)$

$v(F)$

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

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



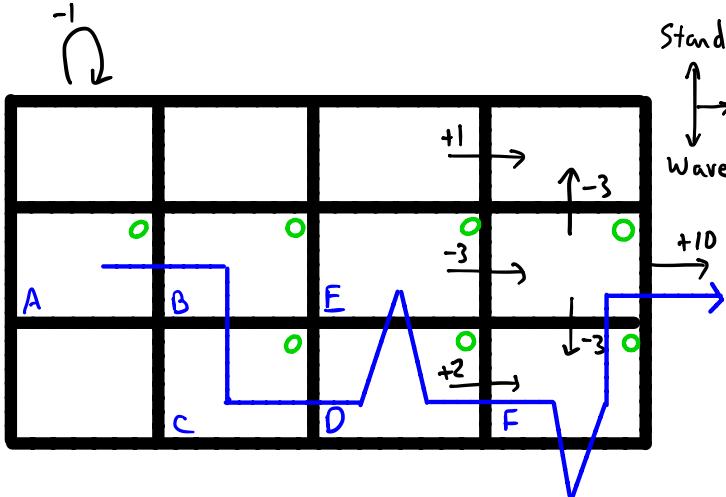
$$\begin{array}{ccccccc}
 & z & & & & & \delta \\
 A & B & C & D & E & F & 0 \\
 | & 0 & 0 & 0 & 0 & 0 & : \\
 \gamma \lambda & (\gamma \lambda)^2 & & & & & \\
 \vdots & (\gamma \lambda)^5 & (\gamma \lambda)^6 & <\text{snip}> & (\gamma \lambda)^8 & \gamma \lambda & 0 \\
 & (\gamma \lambda)^9 & (\gamma \lambda)^{10} & & (\gamma \lambda)^{11} & 0 & 2 \\
 ? & \dots & & & & &
 \end{array}$$

$$\begin{array}{l}
 w = v(A) \\
 2\alpha(\delta \lambda)^5 \\
 v(B) \\
 2\alpha(\delta \lambda)^6 \\
 TD(\lambda) \\
 v(F) \\
 0
 \end{array}$$

1-step

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

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



A	B	C	D	E	F	δ
1	0	0	0	0	0	0
$(\delta \lambda)^2$	$(\delta \lambda)^4$	$(\delta \lambda)^5$	$(\delta \lambda)^3$	$(\delta \lambda)^2 + 1$	$\delta \lambda$	\vdots
$(\delta \lambda)^6$	$(\delta \lambda)^5$	$(\delta \lambda)^4$	$(\delta \lambda)^3 + \delta \lambda$	$(\delta \lambda)^2$	0	2
					1	-1

$<$ snap >

$$w = v(A) \quad 2\alpha(\delta \lambda)^5$$

$$v(B) \quad 2\alpha(\delta \lambda)^4$$

$$TD(\lambda)$$

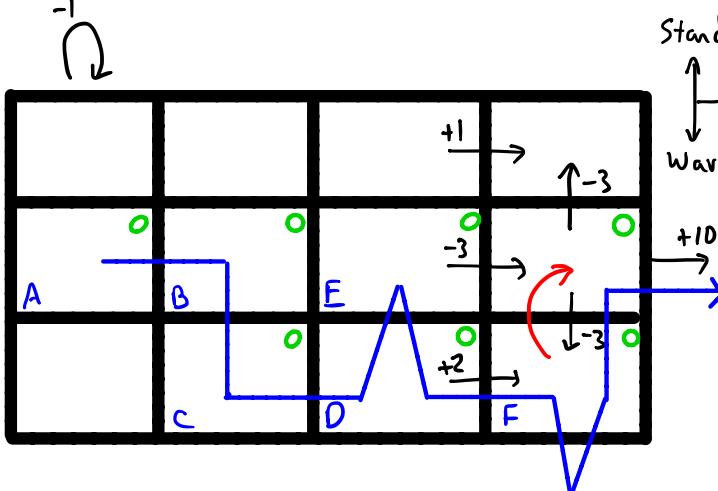
$$v(F) \quad 0 \quad ?$$

1-step

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

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

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



Z						8
A	B	C	D	E	F	O
1	0	0	0	0	0	0
$(\delta \lambda)^3$	$(\delta \lambda)^4$	$(\delta \lambda)^3$	$(\delta \lambda)^2 + 1$	$\delta \lambda$	0	\vdots
$(\delta \lambda)^5$	$(\delta \lambda)^5$	$(\delta \lambda)^4$	$(\delta \lambda)^3 + \delta \lambda$	$(\delta \lambda)^2$	1	2
$(\delta \lambda)^6$	$(\delta \lambda)^5$	$(\delta \lambda)^4$	$(\delta \lambda)^3 + \delta \lambda$	$(\delta \lambda)^2$		-1

$$w = v(A)$$

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

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

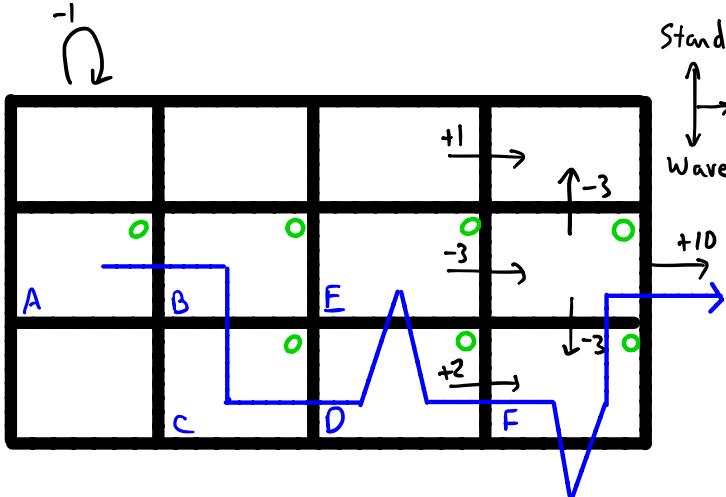
?

TD(2)

$$\begin{matrix} v(F) \\ 0 \\ -\alpha \end{matrix}$$

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

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



A	B	C	D	E	F	δ
1	0	0	0	0	0	0
$(\delta \lambda)^2$	$(\delta \lambda)^4$	$(\delta \lambda)^3$	$(\delta \lambda)^2 + 1$	$\delta \lambda$	0	:
$(\delta \lambda)^5$	$(\delta \lambda)^5$	$(\delta \lambda)^4$	$(\delta \lambda)^3 + \delta \lambda$	$(\delta \lambda)^2$	1	-1
$(\delta \lambda)^6$	$(\delta \lambda)^7$	$(\delta \lambda)^5$	$(\delta \lambda)^2 + 10$	$\delta \lambda^2$	$\delta \lambda^8$	δ

different from
offline case

$$w = v(A)$$

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

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

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

TD(2)

$$v(F)$$

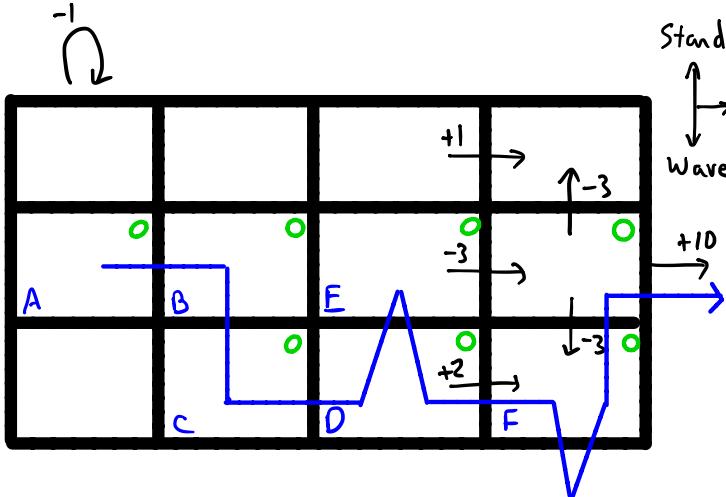
$$0$$

$$-\alpha$$

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

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

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



A	B	C	D	E	F	δ
1	0	0	0	0	0	0
$(\delta \lambda)^2$	$(\delta \lambda)^3$	$(\delta \lambda)^4$	$(\delta \lambda)^5$	$(\delta \lambda)^6$	$(\delta \lambda)^7$	\vdots
$(\delta \lambda)^2$	$(\delta \lambda)^4$	$(\delta \lambda)^6$	$(\delta \lambda)^8$	$(\delta \lambda)^{10}$	$(\delta \lambda)^{12}$	δ
$(\delta \lambda)^3$	$(\delta \lambda)^5$	$(\delta \lambda)^7$	$(\delta \lambda)^9$	$(\delta \lambda)^{11}$	$(\delta \lambda)^{13}$	δ^2
$(\delta \lambda)^4$	$(\delta \lambda)^6$	$(\delta \lambda)^8$	$(\delta \lambda)^{10}$	$(\delta \lambda)^{12}$	$(\delta \lambda)^{14}$	δ^3
$(\delta \lambda)^5$	$(\delta \lambda)^7$	$(\delta \lambda)^9$	$(\delta \lambda)^{11}$	$(\delta \lambda)^{13}$	$(\delta \lambda)^{15}$	δ^4
$(\delta \lambda)^6$	$(\delta \lambda)^8$	$(\delta \lambda)^{10}$	$(\delta \lambda)^{12}$	$(\delta \lambda)^{14}$	$(\delta \lambda)^{16}$	δ^5
$(\delta \lambda)^7$	$(\delta \lambda)^9$	$(\delta \lambda)^{11}$	$(\delta \lambda)^{13}$	$(\delta \lambda)^{15}$	$(\delta \lambda)^{17}$	δ^6
$(\delta \lambda)^8$	$(\delta \lambda)^{10}$	$(\delta \lambda)^{12}$	$(\delta \lambda)^{14}$	$(\delta \lambda)^{16}$	$(\delta \lambda)^{18}$	δ^7
$(\delta \lambda)^9$	$(\delta \lambda)^{11}$	$(\delta \lambda)^{13}$	$(\delta \lambda)^{15}$	$(\delta \lambda)^{17}$	$(\delta \lambda)^{19}$	δ^8
$(\delta \lambda)^{10}$	$(\delta \lambda)^{12}$	$(\delta \lambda)^{14}$	$(\delta \lambda)^{16}$	$(\delta \lambda)^{18}$	$(\delta \lambda)^{20}$	δ^9

different from
 offline case

$$w = v(A)$$

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

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

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

Builds up incrementally to almost same result as offline forward view

$$TD(2)$$

$$v(F)$$

$$0$$

$$-\alpha$$

$$TD(0) \leftrightarrow TD(1)$$

$$TD \dots \lambda \dots MC$$

1-step

A

$$0 \quad (1-\lambda)$$

2

$$0 \quad (1-\lambda)\lambda$$

3

$$0 \quad (1-\lambda)\lambda^2$$

4

$$0 \quad (1-\lambda)\lambda^3$$

5

$$0 \quad (1-\lambda)\lambda^4$$

6

$$2\delta^4 \quad (1-\lambda)\lambda^4$$

7

$$2\delta^5 - \delta^6 \quad (1-\lambda)\lambda^5$$

8

$$2\delta^5 - \delta^6 \quad (1-\lambda)\lambda^6$$

9

$$2\delta^5 - \delta^6 + 10\delta^8 \quad \lambda^8$$

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

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

offline λ -return
 $TD(\lambda)$

offline λ -return

$TD(\lambda)$

(Truncated λ -return)
Online λ -return

True Online $TD(\lambda)$

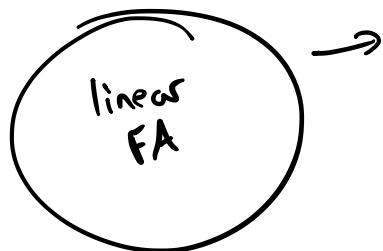
offline λ -return

$TD(\lambda)$

Truncated λ -return

Online λ -return

True Online $TD(\lambda)$



Online λ -return

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

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

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

How differs from TD(λ)?

Online λ -return

$$h = 1 : \quad \mathbf{w}_1^1 \doteq \mathbf{w}_0^1 + \alpha [G_{0:1}^\lambda - \hat{v}(S_0, \mathbf{w}_0^1)] \nabla \hat{v}(S_0, \mathbf{w}_0^1), \quad \leftarrow \text{No change}$$

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

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

How differs from TD(λ)?

Online λ -return

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

\leftarrow No change

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

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

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

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

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

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

How differs from TD(λ)?

Online λ -return

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

\leftarrow No change

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

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

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

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

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

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

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

How differs from TD(λ)?

Online λ -return

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

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

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

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

offline λ -return

$TD(\lambda)$

Truncated λ -return

Online λ -return

True Online $TD(\lambda)$

