Solutions to Reinforcement Learning by Sutton Chapter 12

Yifan Wang

Jan 2020

Exercise 12.1 is too long to be contained in page 1, and is put in the next page. And this is one of the most important exercise in this Chapter as it supports 12.3 and 12.4.

Exercise 12.1

$$
\begin{aligned}
G_t^\lambda &= (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}G_{t:t+n} \\
&= (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\Big[\sum_{k=1}^{n}\gamma^{k-1}R_{t+k} + \gamma^n\hat{v}(S_{t+n},\mathrm{w}_{t+n-1})\Big] \\
&= (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\Big[R_{t+1} + \sum_{k=2}^{n}\gamma^{k-1}R_{t+k} + \gamma^n\hat{v}(S_{t+n},\mathrm{w}_{t+n-1})\Big] \\
&= (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\,R_{t+1} + (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\Big[\sum_{k=2}^{n}\gamma^{k-1}R_{t+k} + \gamma^n\hat{v}(S_{t+n},\mathrm{w}_{t+n-1})\Big] \\
&= (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\,R_{t+1} + (1-\lambda)\sum_{n=1}^{1}\lambda^{n-1}\Big[\sum_{k=2}^{n}\gamma^{k-1}R_{t+k} + \gamma^n\hat{v}(S_{t+n},\mathrm{w}_{t+n-1})\Big] \\
&\quad + \quad \gamma\,(1-\lambda)\sum_{n=2}^{\infty}\lambda^{n-1}\Big[\sum_{k=2}^{n}\gamma^{k-2}R_{t+k} + \gamma^{n-1}\hat{v}(S_{t+n},\mathrm{w}_{t+n-1})\Big] \\
&= (1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\,R_{t+1} + (1-\lambda)\sum_{n=1}^{1}\lambda^{n-1}\Big[\sum_{k=2}^{n}\gamma^{k-1}R_{t+k} + \gamma^n\hat{v}(S_{t+n},\mathrm{w}_{t+n-1})\Big] \\
&\quad + \quad \gamma\,(1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\Big[\sum_{k=2}^{n}\gamma^{k-2}R_{t+k} + \gamma^n\hat{v}(S_{t+n+1},\mathrm{w}_{t+n}) \\
&\quad - \quad \gamma^n\hat{v}(S_{t+n+1},\mathrm{w}_{t+n}) + \gamma^{n-1}\hat{v}(S_{t+n},\mathrm{w}_{t+n-1}) + \gamma^{n-1}R_{t+n+1} - \gamma^{n-1}R_{t+n+1}\Big] - \gamma(1-\lambda)\hat{v}(S_{t+1},\mathrm{w}_t) \\
&= R_{t+1} + \gamma G_{t+1}^\lambda \\
&\quad + \quad \gamma(1-\lambda)\sum_{n=1}^{\infty}\lambda^{n-1}\big[-\gamma^n\hat{v}(S_{t+n+1},\mathrm{w}_{t+n}) + \gamma^{n-1}\hat{v}(S_{t+n},\mathrm{w}_{t+n-1}) - \gamma^{n-1}R_{t+n+1}\big]
\end{aligned}
$$

If you expand the last term you will cancel out a lot of terms and remain the following:

$$
\begin{aligned}
&= R_{t+1} + \gamma G_{t+1}^\lambda + \gamma(1-\lambda)(-G_{t+1}^\lambda) + \gamma(1-\lambda)\hat{v}(S_{t+1},\mathrm{w}_t) \\
&= R_{t+1} + \lambda\gamma G_{t+1}^\lambda + \gamma(1-\lambda)\hat{v}(S_{t+1},\mathrm{w}_t)
\end{aligned}
$$

■

Exercise 12.2

By definition:

$$
\begin{aligned}
(1-\lambda)\lambda^{\tau_\lambda} &= \frac{1}{2}(1-\lambda) \\
\lambda^{\tau_\lambda} &= \frac{1}{2} \\
\tau_\lambda &= \log_\lambda \frac{1}{2}
\end{aligned}
$$

■

Exercise 12.3

Like (6.6), based on a fixed w we form equations as following:

$$
\begin{aligned}
G_t^\lambda - \hat{v}(S_t, \mathrm{w}) &= R_{t+1} + \gamma\lambda G_{t+1}^\lambda - \hat{v}(S_t, \mathrm{w}) + \gamma(1-\lambda)\hat{v}(S_{t+1}, \mathrm{w}) \\
&= R_{t+1} + \gamma\lambda G_{t+1}^\lambda - \hat{v}(S_t, \mathrm{w}) + \gamma\hat{v}(S_{t+1}, \mathrm{w}) - \gamma\hat{v}(S_{t+1}, \mathrm{w}) + \gamma(1-\lambda)\hat{v}(S_{t+1}, \mathrm{w}) \\
&= \delta_t + \gamma\lambda G_{t+1}^\lambda - \gamma\hat{v}(S_{t+1}, \mathrm{w}) + \gamma(1-\lambda)\hat{v}(S_{t+1}, \mathrm{w}) \\
&= \delta_t + \gamma\lambda\big(G_{t+1}^\lambda - \hat{v}(S_{t+1}, \mathrm{w})\big) - \gamma(1-\lambda)\hat{v}(S_{t+1}, \mathrm{w}) + \gamma(1-\lambda)\hat{v}(S_{t+1}, \mathrm{w}) \\
&= \delta_t + \gamma\lambda\Big[\delta_{t+1} + \gamma\lambda\big(G_{t+2}^\lambda - \hat{v}(S_{t+2}, \mathrm{w})\big)\Big] \\
&= \sum_{k=t}^{\infty} \gamma^{k-t}\lambda^{k-t}\delta_k
\end{aligned}
$$

■

Exercise 12.4

$$\sum_{t=0}^{\infty} \alpha \left[(R_t + \gamma \hat{v}(S_{t+1}, \mathrm{w}) - \hat{v}(S_t, w) \right] \mathrm{z}_t \qquad \text{(from 12.7, sum of TD}(\lambda)\text{ updates)}$$

$$= \sum_{t=0}^{\infty} \alpha\, \delta_t \left[\gamma \lambda \mathrm{z}_{t-1} + \nabla \hat{v}(S_t, \mathrm{w}) \right]$$

$$= \sum_{t=0}^{\infty} \alpha\, \delta_t \Big[\gamma \lambda \big(\gamma \lambda z_{t-2} + \nabla \hat{v}(S_{t-1}, \mathrm{w}) \big) + \nabla \hat{v}(S_t, \mathrm{w}) \Big]$$

$$= \sum_{t=0}^{\infty} \alpha\, \delta_t \Big[\gamma^2 \lambda^2 z_{t-2} + \gamma \lambda \nabla \hat{v}(S_{t-1}, \mathrm{w}) + \nabla \hat{v}(S_t, \mathrm{w}) \Big]$$

$$= \sum_{t=0}^{\infty} \alpha\, \delta_t \Big[\sum_{k=0}^{t} \gamma^{t-k} \lambda^{t-k} \nabla \hat{v}(S_k, \mathrm{w}) \Big]$$

Consider for a given t, any $\nabla \hat{v}(S_k, \mathrm{w})$ will only have one occurrence of index in $\sum_{k=0}^{t} \gamma^{t-k} \lambda^{t-k}$ at each $t \geq k$.

Now we collect all those indices across different t's for unique state S_t:

$$= \sum_{t=0}^{\infty} \alpha\, \delta_t \Big[\sum_{k=t}^{\infty} \gamma^{k-t} \lambda^{k-t} \Big] \nabla \hat{v}(S_t, \mathrm{w})$$

Similarly, each δ_t will have indices from all $k \geq t$

$$= \sum_{t=0}^{\infty} \alpha \Big[\sum_{k=t}^{\infty} \gamma^{k-t} \lambda^{k-t} \delta_k \Big] \nabla \hat{v}(S_t, \mathrm{w})$$

$$= \sum_{t=0}^{\infty} \alpha \Big[G_t^{\lambda} - \hat{v}(S_t, \mathrm{w}) \Big] \nabla \hat{v}(S_t, \mathrm{w})$$

(By *exercise 12.3*, and sum of λ-return updates)

■

Exercise 12.5

$$
\begin{aligned}
G^{\lambda}_{t:t+k} &= \sum_{i=1}^{k-1} \lambda^{i-1} G_{t:t+i} - \sum_{i=1}^{k-1} \lambda^{i} G_{t:t+i} + \lambda^{k-1} G_{t:t+k} \qquad (12.9) \\
&= \sum_{i=0}^{k-2} \lambda^{i} G_{t:t+i+1} - \sum_{i=1}^{k-1} \lambda^{i} G_{t:t+i} + \lambda^{k-1} G_{t:t+k} \\
&= G_{t:t+1} - \lambda^{k-1} G_{t:t+k-1} + \sum_{i=1}^{k-2} \lambda^{i} G_{t:t+i+1} - \sum_{i=1}^{k-2} \lambda^{i} G_{t:t+i} + \lambda^{k-1} G_{t:t+k} \\
&= R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathrm{w}_t) + \sum_{i=1}^{k-1} \lambda^{i} \big[G_{t:t+i+1} - G_{t:t+i} \big] \\
&= R_{t+1} + \gamma \hat{v}(S_{t+1}, \mathrm{w}_t) + \sum_{i=t+1}^{t+k-1} (\gamma\lambda)^{i-t} \big[R_{i+1} + \gamma \hat{v}(S_{i+1}, \mathrm{w}_i) - \hat{v}(S_i, \mathrm{w}_{i-1}) \big] \\
&= \hat{v}(S_t, \mathrm{w}_{t-1}) + \sum_{i=t}^{t+k-1} (\gamma\lambda)^{i-t} \big[R_{i+1} + \gamma \hat{v}(S_{i+1}, \mathrm{w}_i) - \hat{v}(S_i, \mathrm{w}_{i-1}) \big] \\
&= \hat{v}(S_t, \mathrm{w}_{t-1}) + \sum_{i=t}^{t+k-1} (\gamma\lambda)^{i-t} \delta'_i
\end{aligned}
$$

■

Exercise 12.6

Replace the loop in $\mathcal{F}(S, A)$ with following:

Loop for i **in** $\mathcal{F}(\mathcal{S}, \mathcal{A})$**:**
| $s \leftarrow s + z_i$
Loop for i **in** $\mathcal{F}(\mathcal{S}, \mathcal{A})$**:**
| $\delta \leftarrow \delta - w_i$
| $z_i \leftarrow \gamma\lambda z_i + 1 - \alpha\gamma\lambda\, s$

Delete the z part in last loop. ■

Exercise 12.7

First, re-do EX 12.1 to check recursive relations under $G^{\lambda s}_{t:h}$. Recall our target $G^{\lambda s}_{t+1:h} \doteq (1-\lambda)\sum_{n=1}^{h-t-2}\lambda^{n-1}G_{t+1:t+1+n} + \lambda^{h-t-2}G_{t+1:h}$.
, where $G_{t+1:t+1+n} = \sum_{k=1}^{n}\gamma^{k-1}R_{t+k+1} + \gamma^n\hat{v}(S_{t+n+1}, \mathrm{w}_{t+n})$

$$
\begin{aligned}
G^{\lambda s}_{t:h} &\doteq (1-\lambda)\sum_{n=1}^{h-t-1}\lambda^{n-1}G_{t:t+n} + \lambda^{h-t-1}G_{t:h} \\
&= (1-\lambda)\sum_{n=1}^{h-t-1}\lambda^{n-1}\Big[\sum_{k=1}^{n}\gamma^{k-1}R_{t+k} + \gamma^n\hat{v}(S_{t+n}, \mathrm{w}_{t+n-1})\Big] + \lambda^{h-t-1}G_{t:h} \\
&= (1-\lambda)\sum_{n=1}^{h-t-1}\lambda^{n-1}\Big[R_{t+1} + \sum_{k=2}^{n}\gamma^{k-1}R_{t+k} + \gamma^n\hat{v}(S_{t+n}, \mathrm{w}_{t+n-1})\Big] + \lambda^{h-t-1}G_{t:h} \\
&= (1-\lambda)\sum_{n=1}^{h-t-1}\lambda^{n-1}\Big[R_{t+1} + \sum_{k=1}^{n-1}\gamma^{k}R_{t+k+1} + \gamma^n\hat{v}(S_{t+n}, \mathrm{w}_{t+n-1})\Big] + \lambda^{h-t-1}G_{t:h} \\
&= (1-\lambda)\sum_{n=1}^{h-t-1}\lambda^{n-1}\Big[R_{t+1} + \gamma\sum_{k=1}^{n-1}\gamma^{k-1}R_{t+k+1} + \gamma^n\hat{v}(S_{t+n}, \mathrm{w}_{t+n-1})\Big] + \lambda^{h-t-1}G_{t:h} \\
&= (1-\lambda)\sum_{n=1}^{h-t-1}\lambda^{n-1}\Big[R_{t+1} + \gamma\sum_{k=1}^{n}\gamma^{k-1}R_{t+k+1} - \gamma\cdot\gamma^{n-1}R_{t+n+1} \\
&\quad + \quad \gamma^{n+1}\hat{v}(S_{t+n+1}, \mathrm{w}_{t+n}) - \gamma^{n+1}\hat{v}(S_{t+n+1}, \mathrm{w}_{t+n}) + \gamma^n\hat{v}(S_{t+n}, \mathrm{w}_{t+n-1})\Big] + \lambda^{h-t-1}G_{t:h} \\
&= (1-\lambda)\sum_{n=1}^{h-t-1}\lambda^{n-1}\Big[R_{t+1} + \gamma G_{t+1:t+n+1} - \gamma^n R_{t+n+1} \\
&\quad - \quad \gamma^{n+1}\hat{v}(S_{t+n+1}, \mathrm{w}_{t+n}) + \gamma^n\hat{v}(S_{t+n}, \mathrm{w}_{t+n-1})\Big] + \lambda^{h-t-1}G_{t:h} \\
&= (1-\lambda)\sum_{n=1}^{h-t-2}\lambda^{n-1}\Big[R_{t+1} + \gamma G_{t+1:t+n+1} - \gamma^n R_{t+n+1} - \gamma^{n+1}\hat{v}(S_{t+n+1}, \mathrm{w}_{t+n}) + \gamma^n\hat{v}(S_{t+n}, \mathrm{w}_{t+n-1})\Big] \\
&+ (1-\lambda)\lambda^{h-t-2}\Big[R_{t+1} + \gamma G_{t+1:h} - \gamma^{h-t-1}R_h - \gamma^{h-t}\hat{v}(S_h, w_{h-1}) + \gamma^{h-t-1}\hat{v}(S_{h-1}, \mathrm{w}_{h-2}))\Big] + \lambda^{h-t-1}G_{t:h} \\
&= \text{(Not completed)}
\end{aligned}
$$

Due to requests from original author of the book, future work are hidden on purpose. Please contact ericwang.usa@gmail.com to refer the answers.