Dhawal Gupta

Extra Parts

1 Proofs

1.1 Proof of $\mathrm{v}_\pi(s)$

$$
\begin{aligned}
\mathrm{v}_\pi(s) &= \mathbb{E}_\pi[G_t|S_t = s] \\
&= \mathbb{E}_\pi[R_{t+1} + \gamma G_{t+1}|S_t = s] \\
&= \mathbb{E}_\pi[R_{t+1}|S_t = s] + \gamma\,\mathbb{E}_\pi[G_{t+1}|S_t = s] \\
&= \sum_{r,a} r \times \pi(a|s) \times p(r|s,a) + \gamma\,\mathbb{E}_\pi[G_{t+1}|S_t = s] \\
&= \sum_{r,a} r\pi(a|s)p(r|s,a) + \\
&\gamma(\sum_{s'}(\sum_{a} \pi(a|s)p(s'|s,a)\,\mathbb{E}_\pi[G_{t+1}|S_{t+1} = s'])) \\
&= \sum_{a} \pi(a|s)[\sum_{r} rp(r|s,a) + \\
&\gamma\sum_{s'} p(s'|s,a)\,\mathbb{E}_\pi[G_{t+1}|S_{t+1} = s']] \\
&= \sum_{a} \pi(a|s)[\sum_{r,s'} rp(s',r|s,a) + \\
&\gamma\sum_{s',r} p(s',r|s,a)\,\mathbb{E}_\pi[G_{t+1}|S_{t+1} = s']] \\
&= \sum_{a} \pi(a|s)\sum_{r,s'} p(s',r|s,a)[r + \gamma\,\mathbb{E}_\pi[G_{t+1}|S_{t+1} = s']] \\
&= \sum_{a} \pi(a|s)\sum_{s',r} p(s',r|s,a)[r + \gamma\,\mathrm{v}_\pi(s')] \forall s \in \mathcal{S}
\end{aligned}
$$

1.2 Proof of Policy Improvement

$$
\begin{aligned}
\mathrm{v}_\pi(s) &\le \mathrm{q}_\pi(s, \pi'(s)) \\
&= \mathbb{E}_{R_{t+1}, S_{t+1} \sim p(s', r|s,a)}[R_{t+1} + \gamma \, \mathrm{v}_\pi(S_{t+1}) | S_t = s, A_t = \pi'(s)] \\
&= \mathbb{E}_{R_{t+1}, S_{t+1} \sim p(s', r|s,a), a \sim \pi'}[R_{t+1} + \gamma \, \mathrm{v}_\pi(S_{t+1}) | S_t = s] \\
&\text{Shortening the notation} \\
&= \mathbb{E}_{\pi'}[R_{t+1} + \gamma \, \mathrm{v}_\pi(S_{t+1}) | S_t = s] \\
&\text{Applying } \pi' \text{ on another step we get} \\
&\le \mathbb{E}_{\pi'}[R_{t+1} + \gamma \, \mathrm{q}_\pi(S_{t+1}, \pi'(S_{t+1}) | S_t = s] \\
&= \mathbb{E}_{\pi'}[R_{t+1} + \gamma \, \mathbb{E}_{\pi'}[R_{t+2} + \gamma \, \mathrm{v}_\pi(S_{t+2}) | S_{t+1}] | S_t = s] \\
&= \mathbb{E}_{\pi'}[R_{t+1} | S_t = s] + \gamma \, \mathbb{E}_{\pi'}[\mathbb{E}_{\pi'}[R_{t+2} + \gamma \, \mathrm{v}_\pi(S_{t+2}) | S_{t+1}] | S_t = s] \\
&= \sum_a \pi'(a|s) \sum_r p(r|s,a) r + \gamma \sum_a \pi'(a|s) \sum_{s'} p(s'|s,a) \, \mathbb{E}_{\pi'}[R_{t+2} + \gamma \, \mathrm{v}_\pi(S_{t+2}) | S_{t+1} = s'] \\
&= \sum_a \pi'(a|s) \sum_{s',r} p(s',r|s,a) r + \gamma \sum_a \pi'(a|s) \sum_{s',r} p(s',r|s,a) \, \mathbb{E}_{\pi'}[R_{t+2} + \gamma \, \mathrm{v}_\pi(S_{t+2}) | S_{t+1} = s'] \\
&= \sum_a \pi'(a|s) \sum_{s',r} p(s',r|s,a)(r + \gamma \, \mathbb{E}_{\pi'}[R_{t+2} + \gamma \, \mathrm{v}_\pi(S_{t+2}) | S_{t+1} = s']) \\
&\text{expanding the inside expectation just like the outside} \\
&= \sum_a \pi'(a|s) \sum_{s',r} p(s',r|s,a)(r + \gamma \sum_a \pi'(a'|s') \sum_{s'',r'} p(s'',r'|s',a')(r' + \gamma \, \mathrm{v}_\pi(s''))) \\
&\text{as we know } \sum_{a'} \pi'(a'|s') \sum_{s'',r'} p(s'',r'|s',a') = \sum_{a',s'',r'} p_{\pi'}(a',r',s''|s') = 1 \\
&\text{and not dependent on } r \\
&= \sum_a \pi'(a|s) \sum_{s',r} p(s',r|s,a)(\sum_{a'} \pi'(a'|s') \sum_{s'',r'} p(s'',r'|s',a') r + \\
&\qquad \gamma \sum_{a'} \pi'(a'|s') \sum_{s'',r'} p(s'',r'|s',a')(r' + \gamma \, \mathrm{v}_\pi(s''))) \\
&= \sum_a \pi'(a|s) \sum_{s',r} p(s',r|s,a) \sum_{a'} \pi'(a'|s') \sum_{s'',r'} p(s'',r'|s',a')(r + \gamma(r' + \gamma \, \mathrm{v}_\pi(s''))) \\
&= \sum_a \pi'(a|s) \sum_{s',r} p(s',r|s,a) \sum_{a'} \pi'(a'|s') \sum_{s'',r'} p(s'',r'|s',a')(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s''))) \\
&= \sum_{a,s',r} p_{\pi'}(s',r,a|s) \sum_{a',s'',r'} p_{\pi'}(s'',r',a'|s')(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s''))) \\
&\text{as internal expression is not dependent on } a \text{ and } a' \text{ we can sum on them}
\end{aligned}
$$

$$= \sum_{s',r} p_{\pi'}(s', r|s) \sum_{s'',r'} p_{\pi'}(s'', r'|s')(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s'')))$$

$$= \sum_{r,s',r',s''} p_{\pi'}(s', r|s) p_{\pi'}(s'', r'|s')(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s'')))$$

we can write this in terms of a, b, c, d, e for ease of readbaility

$= p(a, b|c)p(d, e|a) = p(a|b, c)p(b|c)p(d, e|a)$

$$= \sum_{r,s',r',s''} p_{\pi'}(s'|r, s) p_{\pi'}(r|s) p_{\pi'}(s'', r'|s')(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s'')))$$

we can write $p_{\pi'}(s'', r'|s') = p_{\pi'}(s'', r'|s', s, r)$

as its markov they are equal as they are not dependent on s and r and only on s'

$$= \sum_{r,s',r',s''} p_{\pi'}(s'|r, s) p_{\pi'}(r|s) p_{\pi'}(s'', r'|s', s, r)(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s'')))$$

$$= \sum_{r,s',r',s''} p_{\pi'}(r|s) p_{\pi'}(s'', r', s'|s, r)(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s'')))$$

$$= \sum_{r,s',r',s''} p_{\pi'}(s'', r', s', r|s)(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s'')))$$

we can sum on s' as the independent expression doesnt containt hat term

$$= \sum_{r,r',s''} p_{\pi'}(s'', r', r|s)(r + \gamma r' + \gamma^2 \, \mathrm{v}_\pi(s'')))$$

$$= \mathbb{E}_{\pi'}[R_{t+1} + \gamma R_{t+2} + \gamma^2 \, \mathrm{v}_\pi(S_{t+2})|S_t = s]$$