A Derivation of the Quadratic Form

Proof of Lemma 3. Consider function $h : \mathbb{R} \to \mathbb{R}$,

$$h(w) = c^\top \widehat{V}_{\widetilde{\pi}} w + v_{\widehat{\pi},c}^\top (L_{\pi_w}(\widehat{V}_{\widetilde{\pi}} w) - \widehat{V}_{\widetilde{\pi}} w) \ .$$

For a scalar w, define $\widehat{Q}_{\widetilde{\pi}}(x,a,w) = r(x) + \gamma w(P_{(x,a)} \widehat{V}_{\widetilde{\pi}})$. Substituting for the Bellman operator L_{π_w} (see Section 1.1), we obtain

$$h(w) = c^\top \widehat{V}_{\widetilde{\pi}} w - v_{\widehat{\pi},c}^\top \widehat{V}_{\widetilde{\pi}} w + \sum_x v_{\widehat{\pi},c}(x) \sum_a \nu(a|x) \left(1 + \widehat{Q}_{\widetilde{\pi}}(x,a,w) - \mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w)\right) \widehat{Q}_{\widetilde{\pi}}(x,a,w) \ .$$

Because $\widehat{Q}_{\widetilde{\pi}}(x,a,w) = r(x) + \gamma w P_{(x,a)} \widehat{V}_{\widetilde{\pi}}$, h is quadratic in w, so we can write it as $h(w) = (1/2) w^\top B w + g^\top w + f$ for some choice of parameters B, g, and f. We have that

$$\begin{aligned}
\mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w) &= \sum_a \nu(a|x) \widehat{Q}_{\widetilde{\pi}}(x,a,w) \\
&= \sum_a \nu(a|x)(r(x) + \gamma w P_{(x,a)} \widehat{V}_{\widetilde{\pi}}) \\
&= r(x) + \gamma w \mathbf{E}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}}) \ .
\end{aligned}$$

Also, we have

$$\begin{aligned}
\mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}^2(x,.,w) &= \sum_a \nu(a|x)(\widehat{Q}_{\widetilde{\pi}}(x,a,w))^2 \\
&= \sum_a \nu(a|x)(r(x) + \gamma w P_{(x,a)} \widehat{V}_{\widetilde{\pi}})^2 \\
&= \sum_a \nu(a|x) \left(r(x)^2 + \gamma^2 w^2 (P_{(x,a)} \widehat{V}_{\widetilde{\pi}})^2 + 2\gamma w r(x) P_{(x,a)} \widehat{V}_{\widetilde{\pi}}\right) \\
&= r(x)^2 + 2\gamma w r(x) \mathbf{E}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}}) + \gamma^2 w^2 \mathbf{E}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}})^2 \ .
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{Var}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w) &= \mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}^2(x,.,w) - (\mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w))^2 \\
&= r(x)^2 + 2\gamma w r(x) \mathbf{E}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}}) + \gamma^2 w^2 \mathbf{E}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}})^2 \\
&\quad - r(x)^2 - \gamma^2 w^2 (\mathbf{E}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}}))^2 - 2\gamma w r(x) \mathbf{E}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}}) \\
&= \gamma^2 w^2 \mathbf{Var}_{\nu(.|x)}(P \widehat{V}_{\widetilde{\pi}}) \ .
\end{aligned}$$

Further, we have that

$$\begin{aligned}
h(w) - c^\top \widehat{V}_{\widetilde{\pi}} w + v_{\widehat{\pi},c}^\top \widehat{V}_{\widetilde{\pi}} w &= \sum_x v_{\widehat{\pi},c}(x) \sum_a \nu(a|x) \left(1 + \widehat{Q}_{\widetilde{\pi}}(x,a,w) - \mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w)\right) \widehat{Q}_{\widetilde{\pi}}(x,a,w) \\
&= \sum_x v_{\widehat{\pi},c}(x) \sum_a \nu(a|x) \left(\widehat{Q}_{\widetilde{\pi}}(x,a,w) + (\widehat{Q}_{\widetilde{\pi}}(x,a,w))^2 - \widehat{Q}_{\widetilde{\pi}}(x,a,w) \mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w)\right) \\
&= \sum_x v_{\widehat{\pi},c}(x) \left(\mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w) + \mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w)^2 - (\mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w))^2\right) \\
&= \sum_x v_{\widehat{\pi},c}(x) \mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w) + \sum_x v_{\widehat{\pi},c}(x) \mathbf{Var}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w) \ ,
\end{aligned}$$

and therefore

$$h(w) = c^\top \widehat{V}_{\widetilde{\pi}} w + \sum_x v_{\widehat{\pi},c}(x) \mathbf{E}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w) + \sum_x v_{\widehat{\pi},c}(x) \mathbf{Var}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.,w) - v_{\widehat{\pi},c}^\top \widehat{V}_{\widetilde{\pi}} w \ ,$$

or alternatively,

$$h(w) = v_{\widehat{\pi},c}^\top r + (c^\top \widehat{V}_{\widetilde{\pi}} - v_{\widehat{\pi},c}^\top \widehat{V}_{\widetilde{\pi}} + \gamma \mathbf{E}_{v_{\widehat{\pi},c}}(P^\nu \widehat{V}_{\widetilde{\pi}})) w + w^2 \sum_x v_{\widehat{\pi},c}(x) \mathbf{Var}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x,.) \ .$$

We therefore obtain

$$
\begin{aligned}
f &= v_{\widehat{\pi},c}^\top r \,, \\
g &= c^\top \widehat{V}_{\widetilde{\pi}} - v_{\widehat{\pi},c}^\top \widehat{V}_{\widetilde{\pi}} + \gamma \mathbf{E}_{v_{\widehat{\pi},c}} (P^\nu \widehat{V}_{\widetilde{\pi}}) \,, \\
B &= 2 \sum_x v_{\widehat{\pi},c}(x) \mathbf{Var}_{\nu(.|x)} \widehat{Q}_{\widetilde{\pi}}(x, .) \,.
\end{aligned}
$$

□