Algorithm 1 The policy mirror descent (PMD) method

Input: initial points π_0 and stepsizes $\eta_k \geq 0$.

for $k = 0, 1, \ldots,$ **do**

$$\pi_{k+1}(\cdot|s) = \underset{p(\cdot|s)\in\Delta_{|\mathcal{A}|}}{\arg\min} \left\{ \eta_k [\langle Q^{\pi_k}(s,\cdot), p(\cdot|s)\rangle + h^p(s)] + D^p_{\pi_k}(s) \right\}, \forall s \in \mathcal{S}. \tag{3.5}$$

end for