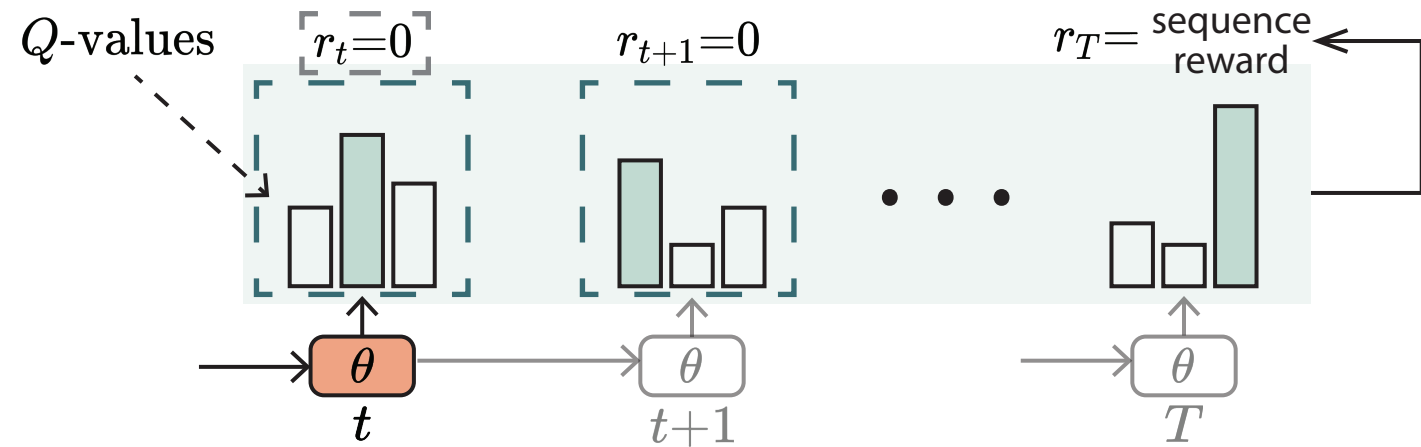
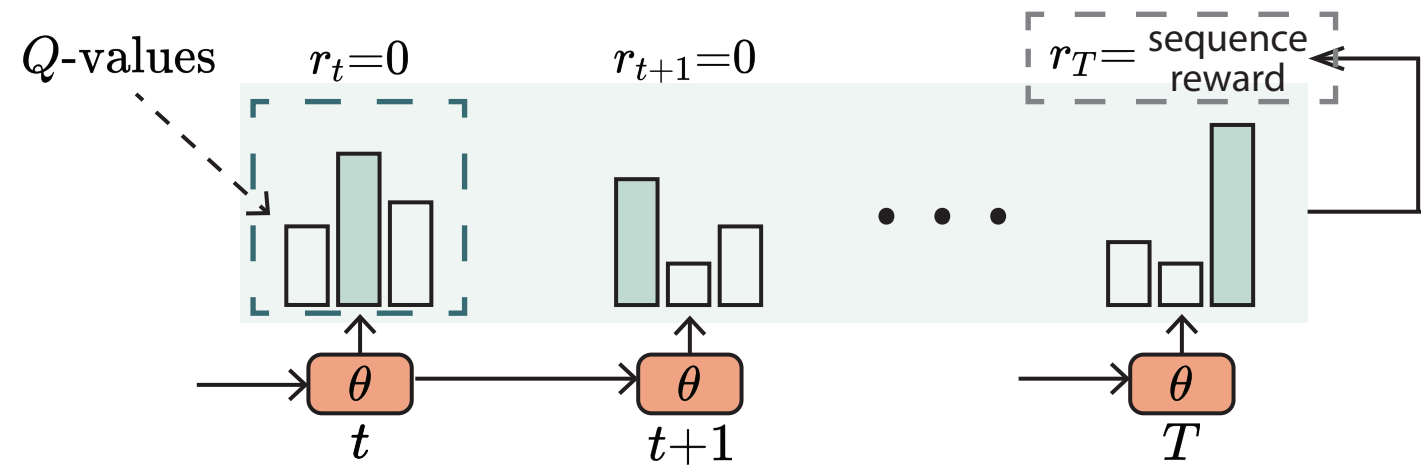


$$\mathcal{L}_{\text{SQL, PCL}}(\boldsymbol{\theta}) = \mathbb{E}_{\pi'} \left[ \frac{1}{2} \left( \begin{array}{c} \boxed{+V_{\bar{\theta}}(\mathbf{s}_t)} \\ \boxed{-} \end{array} + \begin{array}{c} \boxed{+\gamma V_{\bar{\theta}}(\mathbf{s}_{t+1})} \\ \boxed{-} \end{array} + \begin{array}{c} \boxed{+r_t} \\ \boxed{-} \end{array} - \log \pi_{\theta}(a_t \mid \mathbf{s}_t) \right)^2 \right]$$



Single-Step PCL Training

$$\mathcal{L}_{\text{SQL, PCL-ms}}(\boldsymbol{\theta}) = \mathbb{E}_{\pi'} \left[ \frac{1}{2} \left( \begin{array}{c} \boxed{+V_{\bar{\theta}}(\mathbf{s}_t)} \\ \boxed{-} \end{array} + \begin{array}{c} \boxed{+\gamma^{T-t} r_T} \\ \boxed{-} \end{array} + \begin{array}{c} \boxed{+\sum_{l=0}^{T-t} \gamma^l \log \pi_{\theta}(a_{t+l} \mid \mathbf{s}_{t+l})} \\ \boxed{-} \end{array} \right)^2 \right]$$



Multi-Step PCL Training