The optimization problem can be formulated as

$$
\begin{aligned}
&\max_{\pi} \mathbf{E}\left[\sum_{t=0}^{T-1} \gamma^t \mathcal{R}(s_t, a_t)\right], \\
\text{s.t.} &\sum_{t=0}^{T-1}\sum_{k=1}^{K} \mathcal{C}_k(s_t) = 0, \\
&\mathbf{E}\sum_{t=0}^{T-1} \mathcal{G}_j(s_t, a_t) \leq q_j,\ \forall\ 1 \leq j \leq J.
\end{aligned}
\tag{0.1}
$$

If we choose $\gamma \in (0,1)$ sufficiently close to 1 and T not too large, (0.1) is weaker than the following unconstrained optimization problem:

$$
\min_{\substack{\lambda_0 \\ \lambda_j \leq 0, \forall 1 \leq j \leq J}} \max_{\pi} \mathbf{E}\left[\sum_{t=0}^{T-1} \gamma^t \hat{\mathcal{R}}(s_t, a_t)\right]. \tag{0.2}
$$

where

$$
\hat{\mathcal{R}}(s_t, a_t) = \mathcal{R}(s_t, a_t) - \lambda_0 \sum_{k=1}^{K} \mathcal{C}_k(s_t) - \sum_{j=1}^{J} \lambda_j(\mathcal{G}_j(s_t, a_t) - q_j(1-\gamma)^2). \tag{0.3}
$$

Teacher advice by directly modification of the cost function:

1. Value function on both the state and the action space:

$$
\begin{aligned}
\phi_{k+1} = \arg\min_{\phi} \frac{1}{|\mathcal{D}_k| T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T-1} \Big[&(V_\phi(s_t, a_t) - \hat{\mathcal{R}})^2 \\
&+ I_{\{s_t \in S_{adv}\}} \mathrm{ReLU}(V_\phi(s_t, n(s_t)) - V_\phi(s_t, p(s_t)) \Big],
\end{aligned}
\tag{0.4}
$$

   where $p(s)$ is the preferred action and $n(s)$ is the non-preferred action at state s.

2. Value function on only the state space:

$$
\begin{aligned}
&\min_{\phi} \frac{1}{|\mathcal{D}_k| T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T-1} (V_\phi(s_t) - \hat{\mathcal{R}})^2 \\
\text{subject to} \quad &V_\phi(s) \geq c, \quad \forall\ s \in S_{adv_p}, \\
&-V_\phi(s) \geq c, \quad \forall\ s \in S_{adv_n},
\end{aligned}
\tag{0.5}
$$

   where S_{adv_p} is the space of preferred states and S_{adv_n} is the space of non-preferred states.