Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] Theoretical claims are supported by theorems and experimental claims are supported empirically.
   (b) Did you describe the limitations of your work? [Yes] See Section 7.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Section 7.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] We discussed them in each Theorem
   (b) Did you include complete proofs of all theoretical results? [Yes] We include these in Supplementary materials.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We provide these in the supplementary materials.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] We describe these details in the supplementary materials.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] See Section 6.
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] See the Supplementary Materials.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Proof of Theorem 2 & 3

Suppose the loss function $\mathcal{L}$ is decomposable over state action pairs, then we can write, Equation (2) as following:

$$\min_{\hat{\pi}} \max_{\check{\pi}} \sum_{t=1}^{T} \Big[\sum_{\substack{\hat{s},\hat{a}\\ \check{s},\check{a}}} \Big[P(\hat{S}_t = \hat{s}, \hat{A}_t = \hat{a} | \hat{\pi}, \Gamma) \mathcal{L}(\hat{s}, \check{s}, \hat{a}, \check{a}) P(\check{S}_t = \check{s}, \check{A}_t = \check{a} | \check{\pi}, \Gamma) \Big] \Big] \tag{12}$$

$$\text{subject to: } \sum_{t=1}^{T} \sum_{\check{s},\check{a}} P_t(\check{S}_t = \check{s}, \check{A}_t = \check{a} | \check{\pi}, \Gamma) \phi(\check{s}, \check{a}) = \tilde{\boldsymbol{\mu}}.$$

By introducing dual variables $\mathbf{w}$ for the feature expectation constraints, the Lagrangian function of Equation (12) is given by:

$$\min_{\hat{\pi}} \max_{\check{\pi}} \min_{\mathbf{w}} \sum_{t=1}^{T} \Big[\sum_{\substack{\hat{s},\hat{a}\\ \check{s},\check{a}}} \Big[P(\hat{S}_t = \hat{s}, \hat{A}_t = \hat{a} | \hat{\pi}, \Gamma) \mathcal{L}(\hat{s}, \check{s}, \hat{a}, \check{a}) P(\check{S}_t = \check{s}, \check{A}_t = \check{a} | \check{\pi}, \Gamma) \Big] \Big] \tag{13}$$

$$+ \mathbf{w} \cdot \Big(\sum_{t=1}^{T} \Big[\sum_{\check{s},\check{a}} P_t(\check{S}_t = \check{s}, \check{A}_t = \check{a} | \check{\pi}, \Gamma) \phi(\check{s}, \check{a}) \Big] - \tilde{\boldsymbol{\mu}} \Big)$$

The optimization in Equation (13) is over $\hat{\pi}$ and $\check{\pi}$. However, the objective function Equation (13), decomposes over the state-action distribution induced by policies $\hat{\pi}$ and $\check{\pi}$. We directly optimize over the marginals:

$$\min_{(p_1, p_2, \ldots, p_T)} \max_{(q_1, q_2, \ldots, q_T)} \min_{\mathbf{w}} \sum_{t=1}^{T} \Big[\sum_{\substack{\hat{s},\hat{a}\\ \check{s},\check{a}}} \Big[p_t(\hat{s}, \hat{a}) \mathcal{L}(\hat{s}, \check{s}, \hat{a}, \check{a}) q_t(\hat{s}, \hat{a}) \Big] \Big] \tag{14}$$

$$+ \mathbf{w} \cdot \Big(\sum_{t=1}^{T} \Big[\sum_{\check{s},\check{a}} q_t(\check{s}, \check{a}) \phi(\check{s}, \check{a}) \Big] - \tilde{\boldsymbol{\mu}} \Big),$$

where $p_t(\hat{s}, \hat{a}) = P(\hat{S}_t = \hat{s}, \hat{A}_t = \hat{a})$ and $q_t(\check{s}, \check{a}) = P(\check{S}_t = \check{s}, \check{A}_t = \check{a})$. This optimization needs to be over valid state-action marginals (marginals induced by a policy). So the following constrained need to be satisfied:

$$\Omega := \sum_{\hat{s},\hat{a}} p_t(\hat{s}, \hat{a}) P(\hat{s}' | \hat{s}, \hat{a}) = \sum_{\hat{a}} p_t(\hat{s}', \hat{a}) \quad \forall \quad t, s'$$

$$\text{Similarly for} \quad q_t, \quad \sum_{\check{s},\check{a}} q_t(\check{s}, \check{a}) P(\check{s}' | \check{s}, \check{a}) = \sum_{\check{a}} q_t(\check{s}', \check{a}) \quad \forall \quad t, s'$$

Since Equation (14) is convex in all variables p_t, q_t, and $\mathbf{w}$, the order of optimization can be changed:

$$\min_{\mathbf{w}} \max_{\mathbf{Q} \in \Omega} \min_{\mathbf{P} \in \Omega} \Big[\sum_{t=1}^{T} \Big[\sum_{\substack{\hat{s},\hat{a}\\ \check{s},\check{a}}} p_t(\hat{s}, \hat{a}) \mathcal{L}(\hat{s}, \check{s}, \hat{a}, \check{a}) q_t(\hat{s}, \hat{a}) + \mathbf{w} \cdot \sum_{\check{s},\check{a}} q_t(\hat{s}, \hat{a}) \phi(\check{s}, \check{a}) \Big] - \mathbf{w} \cdot \tilde{\boldsymbol{\mu}},$$

where $\mathbf{P} = (p_1, p_2, ..., p_T)$ and $\mathbf{Q} = (q_1, q_2, ..., q_T)$. □

The proof for stationary case is similar.