SUPPLEMENTARY MATERIALS: DISTRIBUTIONALLY ROBUST PARTIALLY OBSERVABLE MARKOV DECISION PROCESS WITH MOMENT-BASED AMBIGUITY*

HIDEAKI NAKAO[†], RUIWEI JIANG[‡], AND SIQIAN SHEN[§]

SM1. Relaxation of a-rectangularity. In this section, we investigate a variant of DR-POMDP where we relax the rectangularity condition of the ambiguity set in the actions. So far, we have only considered the setting where the ambiguity set is rectangular in terms of the states in $\mathcal{S}$ and the actions in $\mathcal{A}$. This is known as (s,a)-rectangular set in the literature of [3], who defined the term in the context of robust MDP. Ref. [3] also considered s-rectangular set in robust POMDP, which is only rectangular in terms of the states $\mathcal{S}$. This setting has randomized policy as the optimal policy. We take a similar approach and formulate the Bellman equation:

$$V^t(\boldsymbol{b}) = \max_{\boldsymbol{\phi}\in\Delta(\mathcal{A})} \min_{\mu\in\mathcal{D}} \mathbb{E}_{P\sim\mu}\left[\sum_{a\in\mathcal{A}} \phi_a \sum_{s\in\mathcal{S}} b_s \left(r_{as} + \beta \sum_{z\in\mathcal{Z}} J_z \boldsymbol{p}_{as} V^{t+1}\left(\boldsymbol{f}\left(\boldsymbol{b}, a, \boldsymbol{p}_a, z\right)\right)\right)\right], \tag{SM1.1}$$

where ϕ_a is the probability for selecting action a. We define the ambiguity set to be

$$\tilde{\mathcal{D}}_s = \left\{ \tilde{\mu}_s \begin{pmatrix} \boldsymbol{p}_s \\ \boldsymbol{r}_s \\ \tilde{\boldsymbol{u}}_s \end{pmatrix} \middle| \begin{array}{l} \mathbb{E}_{(\boldsymbol{p}_s, \boldsymbol{r}_s, \tilde{\boldsymbol{u}}_s)\sim\tilde{\mu}_s}\left[F_s\boldsymbol{p}_s + G_s\boldsymbol{r}_s + H_s\tilde{\boldsymbol{u}}_s\right] = \boldsymbol{c}_s, \\ \tilde{\mu}_s\left(\mathcal{X}_s\right) = 1 \end{array} \right\}, \tag{SM1.2}$$

where $\tilde{\boldsymbol{u}}_s \in \mathbb{R}^Q$ is a vector of auxiliary variables, and

$$\mathcal{X}_s = \left\{ \begin{pmatrix} \boldsymbol{p}_s \\ \boldsymbol{r}_s \\ \tilde{\boldsymbol{u}}_s \end{pmatrix} \in \begin{array}{c} \mathbb{R}^{|\mathcal{A}|\times|\mathcal{S}|\times|\mathcal{Z}|} \\ \mathbb{R}^{|\mathcal{A}|} \\ \mathbb{R}^L \end{array} \middle| B_s\boldsymbol{p}_s + C_s r_s + E_s\tilde{\boldsymbol{u}}_s \preceq_{K_s} \boldsymbol{d}_s \right\}. \tag{SM1.3}$$

Here, $F_s \in \mathbb{R}^{k\times(|\mathcal{A}|\times|\mathcal{S}|\times|\mathcal{Z}|)}$, $G_s \in \mathbb{R}^{k\times|\mathcal{A}|}$, $H_s \in \mathbb{R}^{k\times L}$, $\boldsymbol{c}_s \in \mathbb{R}^k$, $B_s \in \mathbb{R}^{\ell\times(|\mathcal{A}|\times|\mathcal{S}|\times|\mathcal{Z}|)}$, $C_s \in \mathbb{R}^{\ell\times|\mathcal{A}|}$, $E_s \in \mathbb{R}^{\ell\times L}$, and $\boldsymbol{d}_s \in \mathbb{R}^\ell$.

The value function is also convex in the form (4.10), since for $t < T$,

$$\begin{aligned} V^t(\boldsymbol{b}) = \max_{\boldsymbol{\phi}\in\Delta(\mathcal{A})} \max_{\substack{\boldsymbol{\alpha}_{az}\in\mathrm{Conv}(\Lambda^{t+1}) \\ \forall a\in\mathcal{A},\ z\in\mathcal{Z}}} \sum_{s\in\mathcal{S}} b_s \min_{(\hat{\boldsymbol{p}}_s, \hat{\boldsymbol{r}}_s, \hat{\tilde{\boldsymbol{u}}}_s)} \quad & \boldsymbol{\phi}^\top \left(\beta \sum_{z\in\mathcal{Z}} \left[\left(\boldsymbol{\alpha}_{az}^\top J_{az}\right)^\top,\ a\in\mathcal{A}\right]^\top \hat{\boldsymbol{p}}_s + \hat{\boldsymbol{r}}_s\right) \\ \text{s.t.} \quad & F_s\hat{\boldsymbol{p}}_s + G_s\hat{\boldsymbol{r}}_s + H_s\hat{\tilde{\boldsymbol{u}}}_s = \boldsymbol{c}_s, \qquad \forall s\in\mathcal{S} \\ & B_s\hat{\boldsymbol{p}}_s + C_s\hat{\boldsymbol{r}}_s + E_s\hat{\tilde{\boldsymbol{u}}}_s \preceq_{K_s} \boldsymbol{d}_s, \quad \forall s\in\mathcal{S} \end{aligned}$$

*

Funding: This work was funded by the U.S. Department of Engineering (DoE) grant **DE-SC0018018** and National Science Foundation (NSF) grant **CMMI-1727618**.

[†]Department of Industrial and Operations Engineering, University of Michigan at Ann Arbor, USA;

[‡]Department of Industrial and Operations Engineering, University of Michigan at Ann Arbor, USA;

[§]Corresponding author; Department of Industrial and Operations Engineering, University of Michigan at Ann Arbor, USA. Email: siqian@umich.edu.

where $J_{az} \in \mathbb{R}^{|\mathcal{S}|\times(|\mathcal{A}|\times|\mathcal{S}|\times|\mathcal{Z}|)}$ is a matrix of zeros and ones that maps $\boldsymbol{p}_s$ to $\boldsymbol{p}_{asz}$. For an exact algorithm, we solve the inner minimization problem for all $\phi \in \Delta(\mathcal{A})$, $\boldsymbol{\alpha}_{az} \in \mathrm{Conv}(\Lambda^{t+1})$, $\forall z \in \mathcal{Z},\ a \in \mathcal{A}$. The optimal objective is used for constructing the set Λ^t, at each time step t.

SM2. General Ambiguity Set. In this section, we provide a general form of the ambiguity set where the mean values are on an affine manifold, and the supports are conic representable. For all $a \in \mathcal{A}$ and $s \in \mathcal{S}$, we define a non-empty ambiguity set

$$\tilde{\mathcal{D}}_{as} = \left\{ \tilde{\mu}_{as} \begin{pmatrix} \boldsymbol{p}_{as} \\ r_{as} \\ \tilde{\boldsymbol{u}}_{as} \end{pmatrix} \middle| \begin{array}{l} \mathbb{E}_{(\boldsymbol{p}_{as}, r_{as}, \tilde{\boldsymbol{u}}_{as})\sim\tilde{\mu}_{as}} \left[F_{as}\boldsymbol{p}_{as} + G_{as} r_{as} + H_{as}\tilde{\boldsymbol{u}}_{as}\right] = \boldsymbol{c}_{as}, \\ \tilde{\mu}_{as}\left(\mathcal{X}_{as}\right) = 1 \end{array} \right\}, \tag{SM2.1}$$

where $\tilde{\boldsymbol{u}}_{as} \in \mathbb{R}^L$ is a vector of auxiliary variables, and a support with a non-empty relative interior

$$\mathcal{X}_{as} = \left\{ \begin{pmatrix} \boldsymbol{p}_{as} \\ r_{as} \\ \tilde{\boldsymbol{u}}_{as} \end{pmatrix} \in \begin{matrix} \mathbb{R}^{|\mathcal{S}|\times|\mathcal{Z}|} \\ \mathbb{R} \\ \mathbb{R}^L \end{matrix} \middle| B_{as}\boldsymbol{p}_{as} + C_{as} r_{as} + E_{as}\tilde{\boldsymbol{u}}_{as} \preceq_{K_{as}} \boldsymbol{d}_{as} \right\}. \tag{SM2.2}$$

Here, $F_{as} \in \mathbb{R}^{k\times(|\mathcal{S}|\times|\mathcal{Z}|)}$, $G_{as} \in \mathbb{R}^{k\times 1}$, $H_{as} \in \mathbb{R}^{k\times L}$, $\boldsymbol{c}_{as} \in \mathbb{R}^k$, $B_{as} \in \mathbb{R}^{\ell\times(|\mathcal{S}|\times|\mathcal{Z}|)}$, $C_{as} \in \mathbb{R}^{\ell\times 1}$, $E_{as} \in \mathbb{R}^{\ell\times L}$, and $\boldsymbol{d}_{as} \in \mathbb{R}^\ell$. The symbol $\preceq_{K_{as}}$ represents a generalized inequality with respect to a proper cone K_{as}. We denote the marginal distribution by $\mu_{as} = \prod_{(\boldsymbol{p}_{as}, r_{as})} \tilde{\mu}_{as}$, and also extend the definition to the ambiguity set so that $\mathcal{D}_{as} = \prod_{(\boldsymbol{p}_{as}, r_{as})} \tilde{\mathcal{D}}_{as} = \bigcup_{\tilde{\mu}_{as}\in\tilde{\mathcal{D}}_{as}} \prod_{(\boldsymbol{p}_{as}, r_{as})} \tilde{\mu}_{as}$. The auxiliary variables $\tilde{\boldsymbol{u}}_{as}$ are used for "lifting" techniques, enabling the representation of nonlinear constraints to linear ones.

SM3. Proofs of Theorems 4.3 and 4.4.
First, we provide a detailed proof for Theorem 4.3 below.

Proof. We show the result by induction. When $t = T$, $V^T(\boldsymbol{b}) = 0$ satisfies (4.10). For $t < T$, the inner problem $Q^t(\boldsymbol{b}, a)$ described in (4.7) becomes

$$\min_{\tilde{\mu}_a \in \mathcal{P}(\tilde{\mathcal{X}}_a)} \quad \mathbb{E}_{(\boldsymbol{p}_a, \tilde{\boldsymbol{u}}_a)\sim\tilde{\mu}_a} \left[\sum_{s\in\mathcal{S}} b_s \left(r_{as} + \beta \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{as} V^{t+1}\left(\boldsymbol{f}(\boldsymbol{b}, a, \boldsymbol{p}_a, z)\right) \right) \right] \tag{SM3.1a}$$

$$\text{s.t.} \quad \mathbb{E}_{(\boldsymbol{p}_a, \tilde{\boldsymbol{u}}_a)\sim\tilde{\mu}_a} \left[\tilde{\boldsymbol{u}}_{as}\right] = \boldsymbol{c}_{as}, \qquad \forall s \in \mathcal{S} \tag{SM3.1b}$$

$$\mathbb{E}_{(\boldsymbol{p}_a, \tilde{\boldsymbol{u}}_a)\sim\tilde{\mu}_a} \left[I\left((\boldsymbol{p}_{as}, \tilde{\boldsymbol{u}}_{as}) \in \tilde{\mathcal{X}}_{as} \right) \right] = 1, \qquad \forall s \in \mathcal{S} \tag{SM3.1c}$$

for all $a \in \mathcal{A}$. Here $I(\cdot)$ is an indicator function, such that if event $\cdot$ is true, it returns value 1 and 0 otherwise. Associating the dual variables $\boldsymbol{\rho}_{as}$ and ω_{as} with constraints (SM3.1b) and (SM3.1c), respectively, we formulate the dual of (SM3.1) as

$$\max_{\boldsymbol{\rho}_a, \boldsymbol{\omega}_a} \quad \sum_{s\in\mathcal{S}} \boldsymbol{c}_{as}^\top \boldsymbol{\rho}_{as} + \sum_{s\in\mathcal{S}} \omega_{as} \tag{SM3.2a}$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{s\in\mathcal{S}} \tilde{\boldsymbol{u}}_{as}^\top \boldsymbol{\rho}_{as} + \sum_{s\in\mathcal{S}} \omega_{as} \\ & \le \sum_{s\in\mathcal{S}} b_s \left(r_{as} + \beta \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{as} V^{t+1}\left(\boldsymbol{f}(\boldsymbol{b}, a, \boldsymbol{p}_a, z)\right) \right) \qquad \forall (\boldsymbol{p}_a, \tilde{\boldsymbol{u}}_a) \in \tilde{\mathcal{X}}_a \end{aligned} \tag{SM3.2b}$$

$$\boldsymbol{\rho}_{as} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{Z}|},\ \omega_{as} \in \mathbb{R} \qquad \forall s \in \mathcal{S}. \tag{SM3.2c}$$

Constraints (SM3.2b) are further equivalent to the following inequality with a minimization problem on the right-hand side (RHS).

$$\sum_{s\in\mathcal{S}}\omega_{as} \leq \tag{SM3.3a}$$

$$\begin{aligned}
\min_{(\boldsymbol{p}_a,\tilde{\boldsymbol{u}}_a)} \quad & \sum_{s\in\mathcal{S}} b_s\left(r_{as}+\beta\sum_{z\in\mathcal{Z}}\mathbf{1}^\top \boldsymbol{J}_z\boldsymbol{p}_{as}V^{t+1}\left(\boldsymbol{f}(\boldsymbol{b},a,\boldsymbol{p}_a,z)\right)\right)-\sum_{s\in\mathcal{S}}\tilde{\boldsymbol{u}}_{as}^\top\boldsymbol{\rho}_{as} \\
\text{s.t.} \quad & \tilde{\boldsymbol{u}}_{as}\geq \boldsymbol{p}_{as}-\bar{\boldsymbol{p}}_{as} && \forall s\in\mathcal{S} && \text{(SM3.3b)}\\
& \tilde{\boldsymbol{u}}_{as}\geq \bar{\boldsymbol{p}}_{as}-\boldsymbol{p}_{as} && \forall s\in\mathcal{S} && \text{(SM3.3c)}\\
& \mathbf{1}^\top\boldsymbol{p}_{as}=1 && \forall s\in\mathcal{S} && \text{(SM3.3d)}\\
& \boldsymbol{p}_{as}\geq 0 && \forall s\in\mathcal{S}. && \text{(SM3.3e)}
\end{aligned}$$

Substituting (4.10) for V^{t+1} and (4.1) for $\boldsymbol{f}(\boldsymbol{b},a,\boldsymbol{p}_a,z)$, we obtain

$$\begin{aligned}
\text{RHS of (SM3.3)} = \min_{(\boldsymbol{p}_a,\tilde{\boldsymbol{u}}_a)} \quad & \sum_{s\in\mathcal{S}} b_s r_{as}+\beta\sum_{z\in\mathcal{Z}}\max_{\boldsymbol{\alpha}_{az}\in\Lambda^{t+1}}\left[\boldsymbol{\alpha}_{az}^\top\sum_{s\in\mathcal{S}}\boldsymbol{J}_z\boldsymbol{p}_{as}b_s\right]-\sum_{s\in\mathcal{S}}\tilde{\boldsymbol{u}}_{as}^\top\boldsymbol{\rho}_{as}\\
\text{s.t.} \quad & \text{(SM3.3b)–(SM3.3e)}.
\end{aligned} \tag{SM3.4}$$

Since the objective of the maximization problem is linear in terms of $\boldsymbol{\alpha}_{az},\forall z\in\mathcal{Z}$, the optimal objective value does not change by taking the convex hull of Λ^{t+1}, denoted as $\mathrm{Conv}\left(\Lambda^{t+1}\right)$. Bringing the maximization to the front, we have

$$\begin{aligned}
\text{(SM3.4)} = \min_{(\boldsymbol{p}_a,\tilde{\boldsymbol{u}}_a)} \quad & \max_{\substack{\boldsymbol{\alpha}_{az}\in\mathrm{Conv}(\Lambda^{t+1})\\ \forall z\in\mathcal{Z}}}\left[\sum_{s\in\mathcal{S}} b_s r_{as}+\beta\sum_{z\in\mathcal{Z}}\boldsymbol{\alpha}_{az}^\top\sum_{s\in\mathcal{S}}\boldsymbol{J}_z\boldsymbol{p}_{as}b_s-\sum_{s\in\mathcal{S}}\tilde{\boldsymbol{u}}_{as}^\top\boldsymbol{\rho}_{as}\right]\\
\text{s.t.} \quad & \text{(SM3.3b)–(SM3.3e)}
\end{aligned} \tag{SM3.5}$$

The expression in the bracket is convex (linear) in $(\boldsymbol{p}_a,\tilde{\boldsymbol{u}}_a)$ for fixed $\boldsymbol{\alpha}_{az},\ z\in\mathcal{Z}$, and concave (affine) in $\boldsymbol{\alpha}_{az},\ z\in\mathcal{Z}$ given fixed values of $(\boldsymbol{p}_a,\tilde{\boldsymbol{u}}_a)$. Moreover, (SM3.3b)–(SM3.3e) and $\mathrm{Conv}\left(\Lambda^{t+1}\right)$ are convex sets. The minimax theorem (see, e.g., [2], [1]) ensures that the problem is equivalent to

$$\begin{aligned}
\text{(SM3.5)} = \max_{\substack{\boldsymbol{\alpha}_{az}\in\mathrm{Conv}(\Lambda^{t+1})\\ \forall z\in\mathcal{Z}}}\min_{(\boldsymbol{p}_a,\tilde{\boldsymbol{u}}_a)} \quad & \sum_{s\in\mathcal{S}} b_s r_{as}+\beta\sum_{z\in\mathcal{Z}}\boldsymbol{\alpha}_{az}^\top\sum_{s\in\mathcal{S}}\boldsymbol{J}_z\boldsymbol{p}_{as}b_s-\sum_{s\in\mathcal{S}}\tilde{\boldsymbol{u}}_{as}^\top\boldsymbol{\rho}_{as}\\
\text{s.t.} \quad & \text{(SM3.3b)–(SM3.3e)}
\end{aligned} \tag{SM3.6}$$

We take the dual of the inner minimization by associating dual variables $\boldsymbol{\kappa}_{as}^1$, $\boldsymbol{\kappa}_{as}^2$, σ_{as} with constraints (SM3.3b)–(SM3.3d), respectively. We thus have the following equivalence:

$$\begin{aligned}
\text{(SM3.6)} = \max_{\substack{\boldsymbol{\alpha}_{az}\in\mathrm{Conv}(\Lambda^{t+1})\\ \forall z\in\mathcal{Z}}}\max_{\boldsymbol{\kappa}_a^1,\boldsymbol{\kappa}_a^2,\boldsymbol{\sigma}_a} \quad & \sum_{s\in\mathcal{S}} b_s r_{as}+\sum_{s\in\mathcal{S}}\left(-\bar{p}_{as}^\top\boldsymbol{\kappa}_{as}^1+\bar{p}_{as}^\top\boldsymbol{\kappa}_{as}^2+\sigma_{as}\right) && && \text{(SM3.7a)}\\
\text{s.t.} \quad & \beta b_s\sum_{z\in\mathcal{Z}}\boldsymbol{J}_z{}^\top\boldsymbol{\alpha}_{az}+\boldsymbol{\kappa}_{as}^1-\boldsymbol{\kappa}_{as}^2-\mathbf{1}\sigma_{as}\geq 0, && \forall s\in\mathcal{S} && \text{(SM3.7b)}\\
& \boldsymbol{\kappa}_{as}^1+\boldsymbol{\kappa}_{as}^2+\boldsymbol{\rho}_{as}=0, && \forall s\in\mathcal{S} && \text{(SM3.7c)}\\
& \boldsymbol{\kappa}_{as}^1,\boldsymbol{\kappa}_{as}^2\in\mathbb{R}_+^{|\mathcal{S}|\times|\mathcal{Z}|},\sigma_{as}\in\mathbb{R}, && \forall s\in\mathcal{S}, && \text{(SM3.7d)}
\end{aligned}$$

Due to (SM3.3), we substitute $\sum_{s\in\mathcal{S}}\omega_{as}$ in the objective function (SM3.2a) with (SM3.7). As a result, the value function (4.5) is equivalent to

$$\begin{aligned}
&V^t(\boldsymbol{b}) = \max_{a\in\mathcal{A}} \max_{\substack{\boldsymbol{\alpha}_{az}\in\mathrm{Conv}(\Lambda^{t+1})\\ \forall z\in\mathcal{Z}}} \\
&\qquad \max_{\boldsymbol{\rho}_a,\boldsymbol{\kappa}_a^1,\boldsymbol{\kappa}_a^2,\boldsymbol{\sigma}_a} \quad \sum_{s\in\mathcal{S}} \boldsymbol{c}_{as}^\top\boldsymbol{\rho}_{as} + \sum_{s\in\mathcal{S}} b_s r_{as} + \sum_{s\in\mathcal{S}}\left(-\bar{p}_{as}^\top\boldsymbol{\kappa}_{as}^1 + \bar{p}_{as}^\top\boldsymbol{\kappa}_{as}^2 + \sigma_{as}\right) && \text{(SM3.8a)}\\
&\qquad\qquad \text{s.t.} \quad \text{(SM3.7b)–(SM3.7d)}\\
&\qquad\qquad\qquad \boldsymbol{\rho}_{as}\in\mathbb{R}^{|\mathcal{S}|\times|\mathcal{Z}|}\ \forall s\in\mathcal{S}, && \text{(SM3.8b)}
\end{aligned}$$

and after taking the dual of the most inner maximization problem, we have

$$V^t(\boldsymbol{b}) = \max_{a\in\mathcal{A}} \max_{\substack{\boldsymbol{\alpha}_{az}\in\mathrm{Conv}(\Lambda^{t+1})\\ \forall z\in\mathcal{Z}}} \sum_{s\in\mathcal{S}} b_s \times \Xi(a,\boldsymbol{\alpha}_{az}\ \forall z\in\mathcal{Z}, s), \tag{SM3.9}$$

where

$$\begin{aligned}
\Xi(a,\boldsymbol{\alpha}_{az}\ \forall z\in\mathcal{Z}, s) = \min_{(\boldsymbol{p}_{as},\tilde{\boldsymbol{u}}_{as})} \quad & \beta\sum_{z\in\mathcal{Z}} \boldsymbol{\alpha}_{az}^\top \boldsymbol{J}_z\boldsymbol{p}_{as} + r_{as} && \text{(SM3.10a)}\\
\text{s.t.} \quad & \boldsymbol{c}_{as} \geq \boldsymbol{p}_{as} - \bar{\boldsymbol{p}}_{as} && \text{(SM3.10b)}\\
& \boldsymbol{c}_{as} \geq \bar{\boldsymbol{p}}_{as} - \boldsymbol{p}_{as} && \text{(SM3.10c)}\\
& \boldsymbol{1}^\top\boldsymbol{p}_{as} = 1 && \text{(SM3.10d)}\\
& \boldsymbol{p}_{as} \geq 0. && \text{(SM3.10e)}
\end{aligned}$$

Defining set Λ^t as

$$\left\{ \left(\Xi(a,\boldsymbol{\alpha}_{az}\ \forall z\in\mathcal{Z},s),\ s\in\mathcal{S}\right)^\top \,\middle|\, \begin{array}{c}\forall a\in\mathcal{A},\\ \forall\boldsymbol{\alpha}_{az}\in\mathrm{Conv}\left(\Lambda^{t+1}\right),\ \forall z\in\mathcal{Z}\end{array} \right\},$$

it follows that the above value function in (SM3.9) is of the form (4.10). Furthermore, by induction, this is true for all t. This completes the proof. ∎

The proof of Theorem 4.4 is given as follows.

Proof. Consider two arbitrary value functions V_1 and V_2. Given belief state $\boldsymbol{b}$, let

$$a_i^\star = \arg\max_{a\in\mathcal{A}} \min_{\mu_a\in\tilde{\mathcal{D}}_a} \mathbb{E}_{(\boldsymbol{p}_a,\boldsymbol{r}_a)\sim\mu_a}\left[\sum_{s\in\mathcal{S}} b_s\left(r_{as} + \beta\sum_{z\in\mathcal{Z}} \boldsymbol{1}^\top\boldsymbol{J}_z\boldsymbol{p}_{as}V_i\left(\boldsymbol{f}(\boldsymbol{b},a,\boldsymbol{p}_a,z)\right)\right)\right],$$

for $i=1,2$, and for all actions $a\in\mathcal{A}$, denote

$$\mu_{a,i}^\star = \arg\min_{\mu_a\in\tilde{\mathcal{D}}_a} \mathbb{E}_{(\boldsymbol{p}_a,\boldsymbol{r}_a)\sim\mu_a}\left[\sum_{s\in\mathcal{S}} b_s\left(r_{as} + \beta\sum_{z\in\mathcal{Z}} \boldsymbol{1}^\top\boldsymbol{J}_z\boldsymbol{p}_{as}V_i\left(\boldsymbol{f}(\boldsymbol{b},a,\boldsymbol{p}_a,z)\right)\right)\right]$$

for $i = 1, 2$. First, suppose that $\mathcal{L}V_1(\boldsymbol{b}) \geq \mathcal{L}V_2(\boldsymbol{b})$. Then,

$$
\begin{aligned}
0 &\leq \mathcal{L}V_1(\boldsymbol{b}) - \mathcal{L}V_2(\boldsymbol{b}) \\
&= \mathbb{E}_{(\boldsymbol{p}_{a_1^\star}, \boldsymbol{r}_{a_1^\star}) \sim \mu^\star_{a_1^\star,1}} \left[\sum_{s\in\mathcal{S}} b_s \left(r_{a_1^\star s} + \beta \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_1^\star s} V_1\left(\boldsymbol{f}(\boldsymbol{b}, a_1^\star, \boldsymbol{p}_{a_1^\star}, z)\right)\right)\right] \\
&\quad - \mathbb{E}_{(\boldsymbol{p}_{a_2^\star}, \boldsymbol{r}_{a_2^\star}) \sim \mu^\star_{a_2^\star,2}} \left[\sum_{s\in\mathcal{S}} b_s \left(r_{a_2^\star s} + \beta \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_2^\star s} V_2\left(\boldsymbol{f}(\boldsymbol{b}, a_2^\star, \boldsymbol{p}_{a_2^\star}, z)\right)\right)\right] \\
&\leq \mathbb{E}_{(\boldsymbol{p}_{a_1^\star}, \boldsymbol{r}_{a_1^\star}) \sim \mu^\star_{a_1^\star,2}} \left[\sum_{s\in\mathcal{S}} b_s \left(r_{a_1^\star s} + \beta \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_1^\star s} V_1\left(\boldsymbol{f}(\boldsymbol{b}, a_1^\star, \boldsymbol{p}_{a_1^\star}, z)\right)\right)\right] \\
&\quad - \mathbb{E}_{(\boldsymbol{p}_{a_1^\star}, \boldsymbol{r}_{a_1^\star}) \sim \mu^\star_{a_1^\star,2}} \left[\sum_{s\in\mathcal{S}} b_s \left(r_{a_1^\star s} + \beta \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_1^\star s} V_2\left(\boldsymbol{f}(\boldsymbol{b}, a_1^\star, \boldsymbol{p}_{a_1^\star}, z)\right)\right)\right]
\end{aligned}
$$

$$
= \beta \mathbb{E}_{(\boldsymbol{p}_{a_1^\star}, \boldsymbol{r}_{a_1^\star}) \sim \mu^\star_{a_1^\star,2}} \left[\sum_{s\in\mathcal{S}} b_s \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_1^\star s} \times \left(V_1\left(\boldsymbol{f}(\boldsymbol{b}, a_1^\star, z, \boldsymbol{p}_{a_1^\star})\right) - V_2\left(\boldsymbol{f}(\boldsymbol{b}, a_1^\star, \boldsymbol{p}_{a_1^\star}, z)\right)\right)\right]. \tag{SM3.11}
$$

The inequality follows that we replace the nature's optimal decision $\mu^\star_{a_1^\star,1}$ for V_1 by $\mu^\star_{a_1^\star,2}$, and replace the DM's optimal solution $a_2^\star$ for V_2 by $a_1^\star$. Then, by changing the difference between V_1 and V_2 to the absolute value of the difference, we have

$$
\begin{aligned}
\text{(SM3.11)} &\leq \beta \mathbb{E}_{(\boldsymbol{p}_{a_1^\star}, \boldsymbol{r}_{a_1^\star}) \sim \mu^\star_{a_1^\star,2}} \left[\sum_{s\in\mathcal{S}} b_s \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_1^\star s} \times \left| V_1\left(\boldsymbol{f}(\boldsymbol{b}, a_1^\star, \boldsymbol{p}_{a_1^\star}, z)\right) - V_2\left(\boldsymbol{f}(\boldsymbol{b}, a_1^\star, z, \boldsymbol{p}_{a_1^\star})\right)\right|\right] \\
&\leq \beta \mathbb{E}_{(\boldsymbol{p}_{a_1^\star}, \boldsymbol{r}_{a_1^\star}) \sim \mu^\star_{a_1^\star,2}} \left[\sum_{s\in\mathcal{S}} b_s \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_1^\star s} \sup_{\boldsymbol{b}'\in\Delta(\mathcal{S})} \left| V_1(\boldsymbol{b}') - V_2(\boldsymbol{b}')\right|\right] \\
&= \beta \sup_{\boldsymbol{b}'\in\Delta(\mathcal{S})} \left| V_1(\boldsymbol{b}') - V_2(\boldsymbol{b}')\right|.
\end{aligned}
$$

The second inequality follows that we take the supremum for all belief states $\boldsymbol{b}' \in \Delta(\mathcal{S})$, and the last equality is because $\mathbb{E}_{(\boldsymbol{p}_{a_1^\star}, \boldsymbol{r}_{a_1^\star}) \sim \mu^\star_{a_1^\star,2}} \left[\sum_{s\in\mathcal{S}} b_s \sum_{z\in\mathcal{Z}} \mathbf{1}^\top \boldsymbol{J}_z \boldsymbol{p}_{a_1^\star s}\right] = 1$.

The same result holds for the case where $\mathcal{L}V_1(\boldsymbol{b}) < \mathcal{L}V_2(\boldsymbol{b})$. Thus, for any belief state value $\boldsymbol{b}$, it follows that

$$
\left|\mathcal{L}V_1(\boldsymbol{b}) - \mathcal{L}V_2(\boldsymbol{b})\right| \leq \beta \sup_{\boldsymbol{b}'\in\Delta(\mathcal{S})} \left| V_1(\boldsymbol{b}') - V_2(\boldsymbol{b}')\right|,
$$

and therefore,

$$
\sup_{\boldsymbol{b}\in\Delta(\mathcal{S})} \left|\mathcal{L}V_1(\boldsymbol{b}) - \mathcal{L}V_2(\boldsymbol{b})\right| \leq \beta \sup_{\boldsymbol{b}'\in\Delta(\mathcal{S})} \left| V_1(\boldsymbol{b}') - V_2(\boldsymbol{b}')\right|,
$$

yielding that $\mathcal{L}$ is a contraction under $0 < \beta < 1$. This completes the proof. ∎

References.

[1] D.-Z. Du and P. M. Pardalos, *Minimax and Applications*, vol. 4, Springer Science & Business Media, 2013.

[2] T. Osogami, *Robust partially observable Markov decision process*, in International Conference on Machine Learning (ICML), 2015, pp. 106–115.

[3] W. Wiesemann, D. Kuhn, and B. Rustem, *Robust Markov decision processes*, Mathematics of Operations Research, 38 (2013), pp. 153–183.