The Geometry of Robust Value Functions

Kaixin Wang [1] **Navdeep Kumar** [2] **Kuangqi Zhou** [3] **Bryan Hooi** [1 4] **Jiashi Feng** [5] **Shie Mannor** [2 6]

Abstract

The space of value functions is a fundamental concept in reinforcement learning. Characterizing its geometric properties may provide insights for optimization and representation. Existing works mainly focus on the value space for Markov Decision Processes (MDPs). In this paper, we study the geometry of the robust value space for the more general Robust MDPs (RMDPs) setting, where transition uncertainties are considered. Specifically, since we find it hard to directly adapt prior approaches to RMDPs, we start with revisiting the non-robust case, and introduce a new perspective that enables us to characterize both the non-robust and robust value space in a similar fashion. The key of this perspective is to decompose the value space, in a state-wise manner, into unions of hypersurfaces. Through our analysis, we show that the robust value space is determined by a set of *conic hypersurfaces*, each of which contains the robust values of all policies that agree on one state. Furthermore, we find that taking only extreme points in the uncertainty set is sufficient to determine the robust value space. Finally, we discuss some other aspects about the robust value space, including its non-convexity and policy agreement on multiple states.

1. Introduction

The space of value functions for stationary policies is a central concept in Reinforcement Learning (RL), since many RL algorithms are essentially navigating this space to find an optimal policy that maximizes the value function, such as policy gradient (Sutton et al., 1999), policy iteration (Howard, 1960) and evolutionary strategies (de Boer et al., 2005). Characterizing the geometric properties for the space of the value function (*i.e.*, the value space) would offer insights for RL research. A recent work (Dadashi et al., 2019) shows that the value space for Markov Decision Processes (MDPs) is a possibly non-convex polytope, which inspires new methods in representation learning in RL (Bellemare et al., 2019; Dabney et al., 2021).

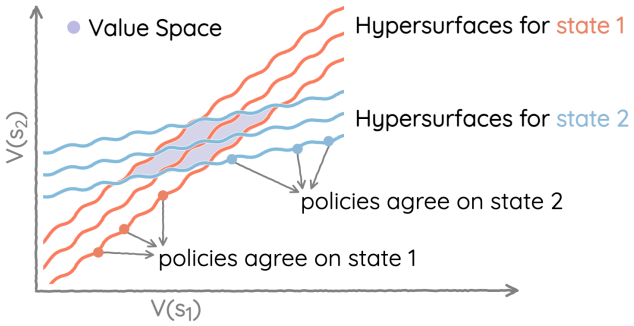

Figure 1. The value space can be decomposed in a state-wise manner as an intersection of unions of hypersurfaces. Each union corresponds to a state and each hypersurface contains the value functions of policies agreeing on that state.

Compared to MDPs, Robust MDPs (RMDPs) are more general, since they do not assume that the transition dynamics are known exactly but instead may take any value from a given uncertainty set (Xu & Mannor, 2006; Iyengar, 2005; Nilim & El Ghaoui, 2005; Wiesemann et al., 2013). This makes RMDPs more suitable for real-world problems where parameters may not be precisely given. Therefore, characterizing the geometric properties of the value space for RMDPs (*i.e.*, robust value space) is of interest.

However, we find it hard to directly adapt the prior approach (Dadashi et al., 2019) from MDPs to RMDPs. Their method builds upon on a key theorem (the Line Theorem), but we find it difficult to prove a robust counterpart of this theorem (see more discussions in Section 5.3).

In this work, we introduce a new perspective for investigating the geometry of the space of value functions. Specifically, we start with revisiting the non-robust case due to its

[1]Institute of Data Science, National University of Singapore, Singapore [2]Electrical and Computer Engineering, Technion, Haifa, Israel [3]Department of Electrical and Computer Engineering, National University of Singapore, Singapore [4]School of Computing, National University of Singapore, Singapore [5]ByteDance, Singapore [6]NVIDIA Research, Haifa, Israel. Correspondence to: Kaixin Wang <kaixin.wang@u.nus.edu>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

simplicity. By decomposing the value space in a state-wise manner (as illustrated in Figure 1), we can give an explicit form about the value function polytope.

With this decomposition-based perspective, we show that the robust value space is determined by a set of *conic hypersurfaces*, each of which contains the robust value functions for policies that agree on one state. Furthermore, from a geometric perspective, we show that the robust value space can be fully determined by a subset of the uncertainty set, which composes of extreme points of the uncertainty set. As a result, for polyhedral uncertainty set such as ℓ_1-ball and ℓ_∞-ball (Ho et al., 2018; 2021; Behzadian et al., 2021), we can replace the infinite uncertainty set with a finite active uncertainty subset, without losing any useful information for policy optimization. Finally, we discuss some other aspects about the robust value space, including policy agreement on more than one state, the non-convexity of the robust value space, and why it is difficult to obtain a Line Theorem for RMDPs.

All proofs and the specifics of MDPs and RMDPs used for illustration can be found in Appendix.

2. Preliminaries

We introduce backgrounds for MDPs in Section 2.1 and for RMDPs in Section 2.2. Importantly, Section 2.3 sets up some essential concepts and notations for studying the value space, which will be frequently used in the rest of paper.

Notations. We use $\mathbf{1}$ and $\mathbf{0}$ to denote vectors of all ones and all zeros respectively, and their sizes can be inferred from the context. For vectors and matrices, $<, \leq, >$ and $\geq$ denote element-wise comparisons. Calligraphic letters such as $\mathcal{P}$ are mainly for sets. For an index set $\mathcal{Z} = \{1, \cdots, k\}$, $(x_i)_{i\in\mathcal{Z}}$ denotes a vector $(x_1, x_2, \cdots, x_k)$ if x_i is a scalar, or a matrix $(x_1, x_2, \cdots, x_k)^\top$ if x_i is a vector. $\Delta_\mathcal{U}$ is used to denote the space of probability distributions over a set $\mathcal{U}$. For a non-empty set $\mathcal{U}$, we denote its *polar cone* as $\mathcal{U}^*$(Bertsekas, 2009), given by

$$\mathcal{U}^* \coloneqq \{y \mid \langle y, x\rangle \leq 0, \forall\, x \in \mathcal{U}\}. \tag{1}$$

We use $\mathbf{conv}(\cdot)$ to denote the convex hull of a set, and $\mathbf{ext}(\cdot)$ to denote the set of extreme points of a non-empty convex set.

2.1. Markov Decision Processes

We consider an MDP $(\mathcal{S}, \mathcal{A}, P, r, \gamma, p_0)$ with a finite state set $\mathcal{S}$ and a finite action set $\mathcal{A}$. The number of states $|\mathcal{S}|$ and the number of actions $|\mathcal{A}|$ are denoted with S and A, respectively. The initial state is generated according to the $p_0 \in \Delta_\mathcal{S}$. We use $P_{s,a} \in \Delta_\mathcal{S}$ to specify the probabilities of transiting to new states when taking action a in state s, and employ $P \coloneqq (P_{s,a})_{s\in\mathcal{S},a\in\mathcal{A}} \in (\Delta_\mathcal{S})^{\mathcal{S}\times\mathcal{A}}$ as a condensed notation. An immediate reward $r_{s,a} \in \mathbb{R}$ is given after taking action a in state s, and similarly $r \coloneqq (r_{s,a})_{s\in\mathcal{S},a\in\mathcal{A}} \in \mathbb{R}^{\mathcal{S}\times\mathcal{A}}$ is a condensed notation. $\gamma \in [0, 1)$ is the discount factor. In addition, we also define $P_s \coloneqq (P_{s,a})_{a\in\mathcal{A}} \in (\Delta_\mathcal{S})^\mathcal{A}$ and $r_s \coloneqq (r_{s,a})_{a\in\mathcal{A}} \in \mathbb{R}^\mathcal{A}$.

A stationary stochastic policy $\pi \coloneqq (\pi_{s,a})_{s\in\mathcal{S},a\in\mathcal{A}} \in (\Delta_\mathcal{A})^\mathcal{S}$ specifies a decision making strategy, where $\pi_{s,a} \in [0, 1]$ is the probability of taking some action a in current state s. We denote $\pi_s \coloneqq (\pi_{s,a})_{a\in\mathcal{A}} \in \Delta_\mathcal{A}$ as the probability vector over actions. In particular, we use $d_{s,a} \in \Delta_\mathcal{A}$ to represent a deterministic π_s that is all-zero except $\pi_{s,a} = 1$.

Under a given policy π, we define the state-to-state transition probability as

$$\begin{aligned} P^\pi &\coloneqq (P^{\pi_s})_{s\in\mathcal{S}} \in (\Delta_\mathcal{S})^\mathcal{S}, \quad \text{where} \\ P^{\pi_s} &\coloneqq P_s\pi_s = \sum_{a\in\mathcal{A}} \pi_{s,a} P_{s,a} \in \Delta_\mathcal{S}. \end{aligned} \tag{2}$$

The reward function under this policy is defined as

$$\begin{aligned} r^\pi &\coloneqq (r^{\pi_s})_{s\in\mathcal{S}} \in \mathbb{R}^\mathcal{S}, \quad \text{where} \\ r^{\pi_s} &\coloneqq r_s^\top \pi_s = \sum_{a\in\mathcal{A}} \pi_{s,a} r_{s,a} \in \mathbb{R}. \end{aligned} \tag{3}$$

The value $V^{\pi,P} \in \mathbb{R}^\mathcal{S}$ is defined to be the expected cumulative reward from starting in a state and acting according to the policy π under transition dynamic P:

$$V^{\pi,P}(s) \coloneqq \mathbb{E}_{P^\pi}\left[\sum_{t=0}^{\infty} \gamma^t r_{s_t,a_t} \mid s_0 = s\right]. \tag{4}$$

2.2. Robust Markov Decision Processes

Robust Markov Decision Processes (RMDPs) generalize MDPs in that the uncertainty in the transition dynamic P is considered (Iyengar, 2005; Nilim & El Ghaoui, 2005; Wiesemann et al., 2013). In an RMDP, the transition dynamic P is chosen adversarially from an uncertainty set $\mathcal{P} \subseteq (\Delta_\mathcal{S})^{\mathcal{S}\times\mathcal{A}}$. We assume throughout the paper that the set $\mathcal{P}$ is compact. The robust value function for a policy π and the optimal robust value function are defined as

$$V^{\pi,\mathcal{P}}(s) \coloneqq \min_{P\in\mathcal{P}} V^{\pi,P}(s), \tag{5}$$

$$V^{\star,\mathcal{P}}(s) \coloneqq \max_{\pi\in\Pi} V^{\pi,\mathcal{P}}(s). \tag{6}$$

Both the policy evaluation and policy improvement problems are intractable for generic $\mathcal{P}$ (Wiesemann et al., 2013). However, they become tractable when certain independence assumptions about $\mathcal{P}$ are made. Two common assumptions are (s, a)-*rectangularity* (Iyengar, 2005; Nilim & El Ghaoui, 2005) and s-*rectangularity* (Wiesemann et al., 2013), which we will use in this paper. The (s, a)-rectangularity assumes

that the adversarial nature selects the worst transition probabilities independently for each state and action. Under (s,a)-rectangularity, the uncertainty set $\mathcal{P}$ can be factorized into $\mathcal{P}_{s,a} \subseteq \Delta_{\mathcal{S}}$ for each state-action pair, *i.e.*,

$$\mathcal{P} = \{P \mid P_{s,a} \in \mathcal{P}_{s,a}, \forall\, s \in \mathcal{S}, \forall\, a \in \mathcal{A}\}, \tag{7}$$

or in short $\mathcal{P} = \times_{(s,a)\in\mathcal{S}\times\mathcal{A}} \mathcal{P}_{s,a}$ where $\times$ denotes Cartesian product. The *s-rectangularity* is less restrictive and assumes the adversarial nature selects the worst transition probabilities independently for each state. Under s-rectangularity, the uncertainty set $\mathcal{P}$ can be factorized into $\mathcal{P}_s \subseteq (\Delta_{\mathcal{S}})^{\mathcal{A}}$ for each state, *i.e.*,

$$\mathcal{P} = \{P \mid P_s \in \mathcal{P}_s, \forall\, s \in \mathcal{S}\}, \tag{8}$$

or in short $\mathcal{P} = \times_{s\in\mathcal{S}} \mathcal{P}_s$. Note that (s,a)-rectangularity is a special case of s-rectangularity. Below we present a restatement of the remark in (Ho et al., 2021) that the optimal policy for the robust policy evaluation MDP is deterministic. This restatement will be used later. Under s-rectangularity, we have for any π,

$$\exists\, P \in \mathcal{P} \quad \text{s.t.} \quad V^{\pi,P}(s) = V^{\pi,\mathcal{P}}(s),\ \forall s \in \mathcal{S}. \tag{9}$$

2.3. The Space of Value Functions

The space of value functions (or value space in short) is the set of value functions for all stationary policies. We use f_P and $f_{\mathcal{P}}$ to respectively represent the mapping between a set of policies and their non-robust and robust value functions, *i.e.*,

$$f_P(U) \coloneqq \{V^{\pi,P} \mid \pi \in U\}, \tag{10}$$

$$f_{\mathcal{P}}(U) \coloneqq \{V^{\pi,\mathcal{P}} \mid \pi \in U\}. \tag{11}$$

The set of all stationary stochastic policies is denoted as $\Pi = (\Delta_{\mathcal{A}})^{\mathcal{S}}$. Then, the non-robust value space for a transition dynamic P and the robust value space for an uncertainty set $\mathcal{P}$ can be respectively expressed as

$$\mathcal{V}^P \coloneqq f_P(\Pi), \tag{12}$$

$$\mathcal{V}^{\mathcal{P}} \coloneqq f_{\mathcal{P}}(\Pi). \tag{13}$$

We then introduce some notations that will be frequently used later. We use Y^{π_s} to denote the set of policies that agree with π on s, *i.e.*,

$$Y^{\pi_s} \coloneqq \{\pi' \mid \pi'_s = \pi_s\}. \tag{14}$$

Note that policy agreement on state s does not imply disagreement on other states. Thus, π itself is also in Y^{π_s}. The row of the matrix $I - \gamma P^\pi$ corresponds to state s is denoted as L^{π_s,P_s}, *i.e.*,

$$L^{\pi_s,P_s} \coloneqq \mathbf{e}_s - \gamma P^{\pi_s} = \mathbf{e}_s - \gamma P_s \pi_s \tag{15}$$

where $\mathbf{e}_s \in \mathbb{R}^{\mathcal{S}}$ is an all-zero vector except the entry corresponding to s being 1.

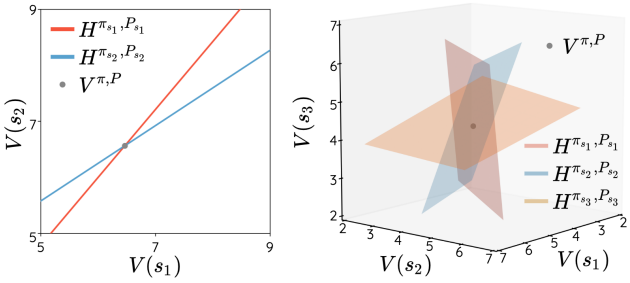

Figure 2. Hyperplanes H^{π_s,P_s} corresponding to different s intersect at the value function $V^{\pi,P}$.

3. The Value Function Polytope Revisited

In this section, we revisit the non-robust value space from a new perspective, where the value space is decomposed in a state-wise manner. This perspective enables us to characterize the polytope shape of the value space in a more straightforward way, and leads to an explicit form of the value polytope.

Our first step is to connect a single value function $V^{\pi,P}$ to a set of hyperplanes, each of which can be expressed as:

$$H^{\pi_s,P_s} \coloneqq \{\mathbf{x} \in \mathbb{R}^{\mathcal{S}} \mid \langle \mathbf{x}, L^{\pi_s,P_s} \rangle = r^{\pi_s}\}. \tag{16}$$

As shown in Lemma 3 in (Dadashi et al., 2019), the value functions $f_P(Y^{\pi_s})$ lie in the hyperplane H^{π_s,P_s}.

Specifically, since $\pi \in Y^{\pi_s}$, we know every hyperplane H^{π_s,P_s} passes through $V^{\pi,P}$ (see examples in Figure 2). The following lemma states that this intersecting point is unique.

Lemma 3.1. *Consider a policy π and a transition dynamic P, we have*

$$\{V^{\pi,P}\} = \bigcap_{s\in\mathcal{S}} H^{\pi_s,P_s} \tag{17}$$

Lemma 3.1 bridges between a single value function and the intersection of S different hyperplanes, each of which corresponds to a state s. Then, by definition (Eqn. (12)), we can obtain the value space by taking the union over all $\pi \in \Pi$, *i.e.*,

$$\mathcal{V}^P = \bigcup_{\pi\in\Pi} \bigcap_{s\in\mathcal{S}} H^{\pi_s,P_s}, \tag{18}$$

as illustrated in Figure 3(a).

From Eqn. (18), we observe that the value space $\mathcal{V}^P$ can also be expressed from an alternative perspective (as shown in Figure 3(b)): 1) for each state $s \in \mathcal{S}$, taking the union of all hyperplanes corresponding to different $\pi_s \in \Delta_{\mathcal{A}}$; 2) taking the intersection of the unions obtained in previous step. The following lemma formalizes this perspective.

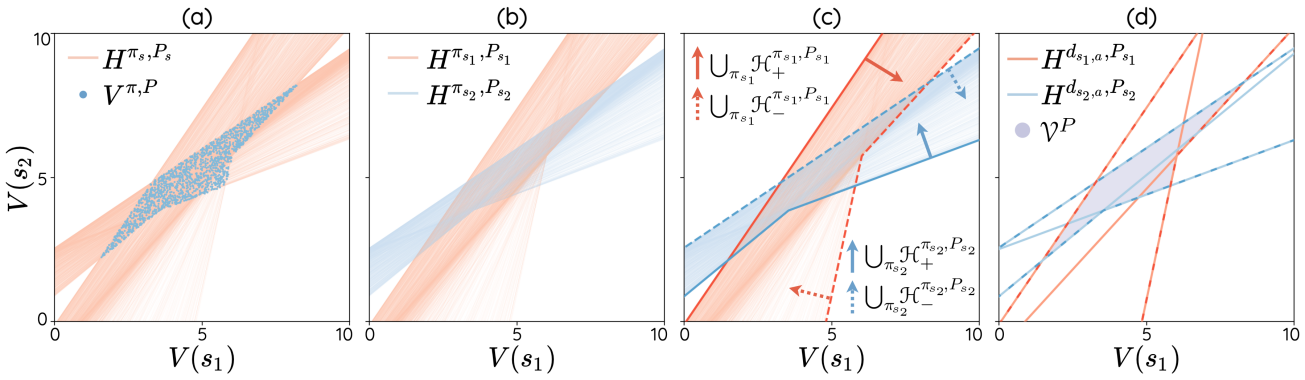

Figure 3. Visualization of the value functions for a 2-state 3-action MDP. **(a)** For each policy π, we plot the value function $V^{\pi,P}$ and the corresponding hyperplanes H^{π_s,P_s} intersecting at $V^{\pi,P}$. **(b)** For each policy π, the hyperplanes H^{π_s,P_s} intersecting at $V^{\pi,P}$ are plotted in different colors for different states. **(c)** For each state s, the union of $\mathcal{H}_+^{\pi_s,P_s}$ and the union of $\mathcal{H}_-^{\pi_s,P_s}$ over all $\pi_s \in \Delta_\mathcal{A}$ are highlighted respectively. **(d)** For each state s, the hyperplanes $H^{d_{s,a},P_s}$ for different actions a are plotted. The union of $\mathcal{H}_+^{d_{s,a},P_s}$ and the union of $\mathcal{H}_-^{d_{s,a},P_s}$ over all actions $a \in \mathcal{A}$ are highlighted as dashed. The entire value space $\mathcal{V}^P$ is visualized as the purple region.

Lemma 3.2. *Consider a transition dynamic P, the value space $\mathcal{V}^P$ can be represented as*

$$\mathcal{V}^P = \bigcup_{\pi\in\Pi}\bigcap_{s\in\mathcal{S}} H^{\pi_s,P_s} = \bigcap_{s\in\mathcal{S}}\bigcup_{\pi_s\in\Delta_\mathcal{A}} H^{\pi_s,P_s}. \tag{19}$$

As suggested in Lemma 3.2, the core of this perspective is to decompose the value space in a state-wise manner. In this way, to study the whole value space, we only need to focus on the union of hyperplanes corresponding to one state.

Specifically, let us denote the two closed half-spaces determined by the hyperplane H^{π_s,P_s} as

$$\begin{aligned}\mathcal{H}_+^{\pi_s,P_s} &:= \{\mathbf{x}\in\mathbb{R}^\mathcal{S} \mid \langle \mathbf{x}, L^{\pi_s,P_s}\rangle \geq r^{\pi_s}\},\\ \mathcal{H}_-^{\pi_s,P_s} &:= \{\mathbf{x}\in\mathbb{R}^\mathcal{S} \mid \langle \mathbf{x}, L^{\pi_s,P_s}\rangle \leq r^{\pi_s}\}.\end{aligned} \tag{20}$$

Then the value space can be expressed in terms of the half-spaces:

$$\mathcal{V}_P = \bigcap_{s\in\mathcal{S}}\bigcup_{\pi_s\in\Delta_\mathcal{A}} \mathcal{H}_+^{\pi_s,P_s}\cap\mathcal{H}_-^{\pi_s,P_s}. \tag{21}$$

Recall that in (Dadashi et al., 2019) a convex polyhedron is defined as a finite intersection of half-spaces, and a polytope is a bounded finite union of convex polyhedra. So our goal is to get rid of this infinite union $\bigcup_{\pi_s\in\Delta_\mathcal{A}}$.

To this end, we first replace the inner union in Eqn. (21) with an intersection of two unions, as illustrated in Figure 3(c) and formally stated in the following lemma.

Lemma 3.3. *Consider a policy π and a transition dynamic P, we have for all states $s \in \mathcal{S}$,*

$$\bigcup_{\pi_s\in\Delta_\mathcal{A}} \mathcal{H}_+^{\pi_s,P_s}\cap\mathcal{H}_-^{\pi_s,P_s} = \left[\bigcup_{\pi_s\in\Delta_\mathcal{A}} \mathcal{H}_+^{\pi_s,P_s}\right]\cap\left[\bigcup_{\pi_s\in\Delta_\mathcal{A}} \mathcal{H}_-^{\pi_s,P_s}\right]. \tag{22}$$

Although these two unions are still taken over infinite set $\Delta_\mathcal{A}$, the following Lemma 3.4 shows that they actually coincide with the finite unions of half-spaces that correspond to $d_{s,a}$ (*i.e.*, deterministic π_s). We can get an intuition by comparing Figure 3(c) and Figure 3(d).

Lemma 3.4. *Consider a policy π and a transition dynamic P, we have for all states $s \in \mathcal{S}$,*

$$\bigcup_{\pi_s\in\Delta_\mathcal{A}} \mathcal{H}_\delta^{\pi_s,P_s} = \bigcup_{a\in\mathcal{A}} \mathcal{H}_\delta^{d_{s,a},P_s}, \quad \forall\,\delta\in\{+,-\}. \tag{23}$$

Finally, putting everything together, we are able to represent the value space with finite union and intersection operations on half-spaces, as stated in Theorem 3.5 and illustrated in Figure 3(d). Using the distributive law of sets, we can see that the value space $\mathcal{V}^P$ immediately satisfies the definition of polyhedron. Since $\mathcal{V}^P$ is bounded, we can conclude that $\mathcal{V}^P$ is a polytope.

Theorem 3.5. *Consider a transition dynamic P, the value space $\mathcal{V}^P$ can be represented as*

$$\begin{aligned}\mathcal{V}^P &= \bigcap_{s\in\mathcal{S}}\left[\left[\bigcup_{a\in\mathcal{A}}\mathcal{H}_+^{d_{s,a},P_s}\right]\cap\left[\bigcup_{a\in\mathcal{A}}\mathcal{H}_-^{d_{s,a},P_s}\right]\right]\\ &= \bigcup_{\mathbf{a}\in\mathcal{A}^\mathcal{S}}\bigcup_{\mathbf{a}'\in\mathcal{A}^\mathcal{S}}\bigcap_{s\in\mathcal{S}}\left[\mathcal{H}_+^{d_{s,\mathbf{a}_s},P_s}\cap\mathcal{H}_-^{d_{s,\mathbf{a}'_s},P_s}\right]\end{aligned} \tag{24}$$

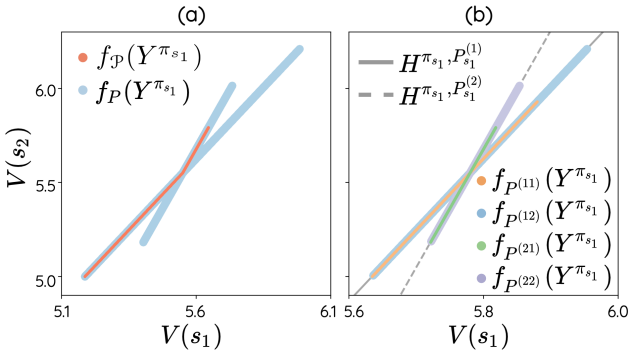

Figure 4. Visualization of the robust value functions for a 2-state 2-action RMDP with an s-rectangular uncertainty set. We consider $\mathcal{P} = \mathcal{P}_{s_1} \times \mathcal{P}_{s_2}$ with $\mathcal{P}_{s_1} = \{P_{s_1}^{(1)}, P_{s_1}^{(2)}\}$ and $\mathcal{P}_{s_2} = \{P_{s_2}^{(1)}, P_{s_2}^{(2)}\}$. Denote $P^{(ij)} \in \mathcal{P}$ such that $P_{s_1}^{(ij)} = P_{s_1}^{(i)}$ and $P_{s_2}^{(ij)} = P_{s_2}^{(j)}$. We plot with different widths to differentiate overlapping lines. **(a)** For the same set of policies $Y^{\pi_{s_1}}$, the set of non-robust value functions $f_P(Y^{\pi_{s_1}})$ for different $P \in \mathcal{P}$ and the set of robust value functions $f_{\mathcal{P}}(Y^{\pi_{s_1}})$ are plotted. **(b)** For different $P \in \mathcal{P}$, $f_P(Y^{\pi_{s_1}})$ are highlighted in different colors. The hyperplanes corresponding to different $P_{s_1} \in \mathcal{P}_{s_1}$ are plotted.

where $\mathbf{a} = (\mathbf{a}_s)_{s\in\mathcal{S}}$, $\mathbf{a}' = (\mathbf{a}'_s)_{s\in\mathcal{S}}$, *and* $\mathbf{a}_s, \mathbf{a}'_s \in \mathcal{A}$.

Compared to the prior approach (Dadashi et al., 2019), our work gives an explicit form of the value function polytope, showing how the value polytope is formed (*cf*. the proof of Proposition 1 in (Dadashi et al., 2019)).

4. Value Space Geometry of RMDPs

4.1. Policy Agreement and the Conic Hypersurface

Recall that in Section 3, our new perspective connects the value space to the hyperplanes where $f_P(Y^{\pi_s})$ lies. Thus in order to characterize the robust value space, we start with studying the geometric properties of robust value functions for all policies that agree on one state, *i.e.*, $f_{\mathcal{P}}(Y^{\pi_s})$. Unlike the non-robust case, $f_{\mathcal{P}}(Y^{\pi_s})$ may not lie in a hyperplane, as shown in Figure 4(a). Nevertheless, it looks like $f_{\mathcal{P}}(Y^{\pi_s})$ still lies in a hypersurface (also see the example for $|\mathcal{S}| = 3$ in the supplementary). In what follows, we are going to characterize this hypersurface.

First, as shown in Figure 4(b), for different $P \in \mathcal{P}$ that share the same P_s, their $f_P(Y^{\pi_s})$ lie in the same hyperplane H^{π_s,P_s}. Comparing Figure 4(a) and (b), it seems that the robust value functions $f_{\mathcal{P}}(Y^{\pi_s})$ always lie in the lower half-space $\mathcal{H}_-^{\pi_s,P_s}$ for different $P \in \mathcal{P}$. On the other hand, from Eqn. (9), we know that there exists $P_s \in \mathcal{P}_s$ such that $V^{\pi,\mathcal{P}}$ lies in the hyperplane H^{π_s,P_s}. Putting it together, we have the following lemma about $f_{\mathcal{P}}(Y^{\pi_s})$.

Lemma 4.1. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have for all states $s \in \mathcal{S}$,*

$$f_{\mathcal{P}}(Y^{\pi_s}) \subseteq \left[\bigcap_{P_s\in\mathcal{P}_s} \mathcal{H}_-^{\pi_s,P_s}\right] \cap \left[\bigcup_{P_s\in\mathcal{P}_s} H^{\pi_s,P_s}\right]. \quad (25)$$

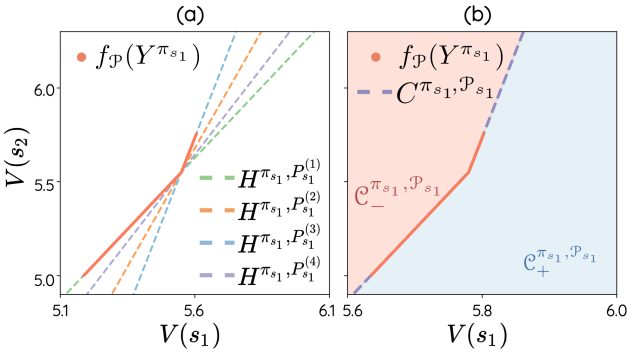

Figure 5. **(a)** For different $P_s \in \mathcal{P}_s$, the hyperplanes H^{π_s,P_s} intersect at one point. **(b)** Illustration of the conic hypersurface in which $f_{\mathcal{P}}(Y^{\pi_s})$ lies.

Note that the right hand side (RHS) of above Eqn. (25) is essentially the boundary of the intersection of half-spaces $\bigcap_{P_s\in\mathcal{P}_s} \mathcal{H}_-^{\pi_s,P_s}$. To further characterize the geometry, we need to know how these half-spaces intersect (equivalently how the hyperplanes intersect). One interesting observation is that when $\mathcal{P}_s$ contains more then 2 elements, the hyperplanes still intersect at one point, as illustrated in Figure 5(a). The following lemma states this property and also gives the intersecting point.

Lemma 4.2. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have for all states $s \in \mathcal{S}$,*

$$\frac{r^{\pi_s}}{1-\gamma}\mathbf{1} \in \bigcap_{P_s\in\mathcal{P}_s} H^{\pi_s,P_s}. \quad (26)$$

Since the hyperplanes intersect at the same point, the intersection of the half-spaces will be a convex cone. We denote

$$\begin{aligned}\mathcal{C}_+^{\pi_s,\mathcal{P}_s} &= \{\mathbf{x} \in \mathbb{R}^{\mathcal{S}} \mid \langle \mathbf{x}, L^{\pi_s,P_s}\rangle \geq r^{\pi_s},\ \exists P_s \in \mathcal{P}_s\},\\ \mathcal{C}_-^{\pi_s,\mathcal{P}_s} &= \{\mathbf{x} \in \mathbb{R}^{\mathcal{S}} \mid \langle \mathbf{x}, L^{\pi_s,P_s}\rangle \leq r^{\pi_s},\ \forall P_s \in \mathcal{P}_s\}.\end{aligned} \quad (27)$$

The following corollary characterizes the hypersurface that $f_{\mathcal{P}}(Y^{\pi_s})$ lies in. Figure 5(b) gives an illustration.

Corollary 4.3. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have for all states $s \in \mathcal{S}$,*

$$f_{\mathcal{P}}(Y^{\pi_s}) \subseteq C^{\pi_s,\mathcal{P}_s} \quad (28)$$

where $C^{\pi_s,\mathcal{P}_s} = \mathcal{C}_+^{\pi_s,\mathcal{P}_s} \cap \mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ is a conic hypersurface.

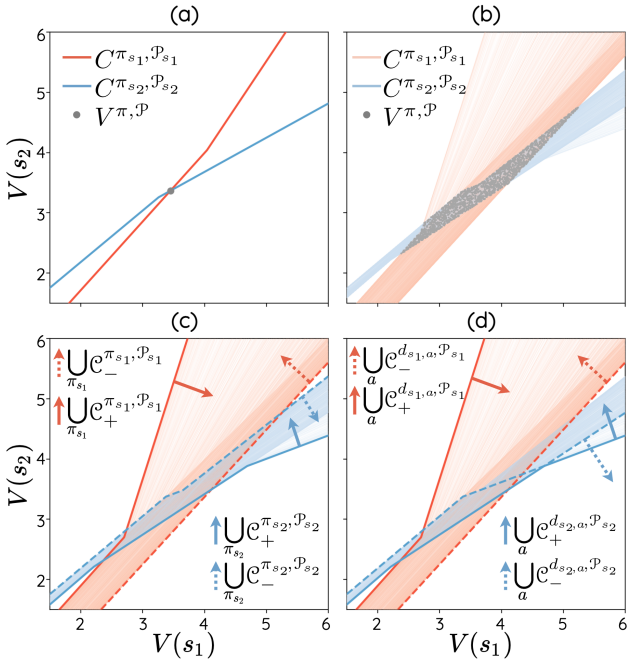

Figure 6. Visualizations of the robust value functions for a 2-state 2-action RMDP with an s-rectangular uncertainty set. **(a)** For a fixed π, the conic hypersurfaces $C^{\pi_s,\mathcal{P}_s}$ corresponding to different s intersect at the robust value function $V^{\pi,\mathcal{P}}$. **(b)** For each policy π, the robust value function $V^{\pi,\mathcal{P}}$ is plotted, and the corresponding conic hypersurfaces $C^{\pi_s,\mathcal{P}_s}$ intersecting at $V^{\pi,\mathcal{P}}$ are plotted in different colors for different states. **(c)** For each state s, the union of $\mathcal{C}_+^{\pi_s,\mathcal{P}_s}$ and the union of $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ over all $\pi_s \in \Delta_{\mathcal{A}}$ are highlighted respectively. **(d)** For each state s, the union of $\mathcal{C}_+^{d_{s,a},\mathcal{P}_s}$ and the union of $\mathcal{C}_-^{d_{s,a},\mathcal{P}_s}$ over all $a \in \mathcal{A}$ are highlighted respectively.

4.2. The Robust Value Space

With the knowledge about the geometry of $f_{\mathcal{P}}(Y^{\pi_s})$, we are now ready to characterize the entire robust value space $\mathcal{V}^{\mathcal{P}}$. Similar to Section 3, we first connect the single robust value function to the intersection of S different conic hypersurfaces by the following lemma (see Figure 6(a) for an illustration).

Lemma 4.4. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have*

$$\{V^{\pi,\mathcal{P}}\} = \bigcap_{s\in\mathcal{S}} C^{\pi_s,\mathcal{P}_s}. \tag{29}$$

Then from the introduced perspective, we show that the robust value space can also be viewed as an intersection of state-wise unions of conic hypersurfaces, as illustrated in Figure 6(b) and formally stated in Lemma 4.5.

Lemma 4.5. *Consider an s-rectangular uncertainty set $\mathcal{P}$,*

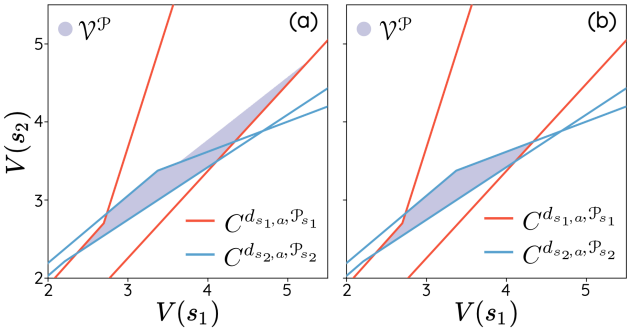

Figure 7. The robust value space $\mathcal{V}^{\mathcal{P}}$ and the conic hypersurfaces $C^{d_{s,a},\mathcal{P}_s}$ under s-rectangularity **(a)** and (s,a)-rectangularity **(b)**.

the robust value function space $\mathcal{V}^{\mathcal{P}}$ can be represented as

$$\mathcal{V}^{\mathcal{P}} = \bigcup_{\pi\in\Pi} \bigcap_{s\in\mathcal{S}} C^{\pi_s,\mathcal{P}_s} = \bigcap_{s\in\mathcal{S}} \bigcup_{\pi_s\in\Delta_{\mathcal{A}}} C^{\pi_s,\mathcal{P}_s}. \tag{30}$$

Next, we show the equivalence between each inner union in RHS of the above equation and an intersection of two unions in Lemma 4.6. Figure 6(c) gives an illustration. Similar to the non-robust case, Lemma 4.6 will help us characterize the relationship between the robust value space $\mathcal{V}^{\mathcal{P}}$ and the conic hypersurfaces corresponding to $d_{s,a}$.

Lemma 4.6. *Consider an s-rectangular uncertainty set $\mathcal{P}$, we have for all states $s \in \mathcal{S}$,*

$$\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} C^{\pi_s,\mathcal{P}_s} = \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_+^{\pi_s,\mathcal{P}_s}\right] \cap \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_-^{\pi_s,\mathcal{P}_s}\right]. \tag{31}$$

As shown in Figure 6(d), unlike the non-robust case, the infinite union $\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ does not necessarily coincides with the finite union $\bigcup_{a\in\mathcal{A}} \mathcal{C}_-^{d_{s,a},\mathcal{P}_s}$. The following Lemma 4.7 characterizes their relationship.

Lemma 4.7. *Consider an s-rectangular uncertainty set $\mathcal{P}$, we have for all states $s \in \mathcal{S}$,*

$$\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_+^{\pi_s,\mathcal{P}_s} = \bigcup_{a\in\mathcal{A}} \mathcal{C}_+^{d_{s,a},\mathcal{P}_s}, \tag{32}$$

$$\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_-^{\pi_s,\mathcal{P}_s} \supseteq \bigcup_{a\in\mathcal{A}} \mathcal{C}_-^{d_{s,a},\mathcal{P}_s}, \tag{33}$$

where the equality in the second line holds when $\mathcal{P}$ is (s,a)-rectangular.

Putting it together, the robust value space can be characterized in Theorem 4.8. Figure 7 highlights the difference in the robust value space between s-rectangularity and (s,a)-rectangularity, by using the same set of probability values

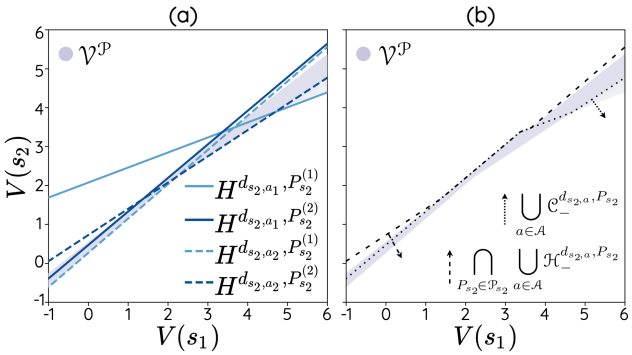

Figure 8. A closer look at the "extra" region under s-rectangularity. Here $\mathcal{P}_{s_2} = \{P_{s_2}^{(1)}, P_{s_2}^{(2)}\}$. We highlight the hyperplanes in **(a)**, and the upper and lower bounds of the region $\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}^{\pi_s,\mathcal{P}_s}$ in **(b)**. Note that the black boundaries in **(b)** are composed by the hyperplanes in **(a)**.

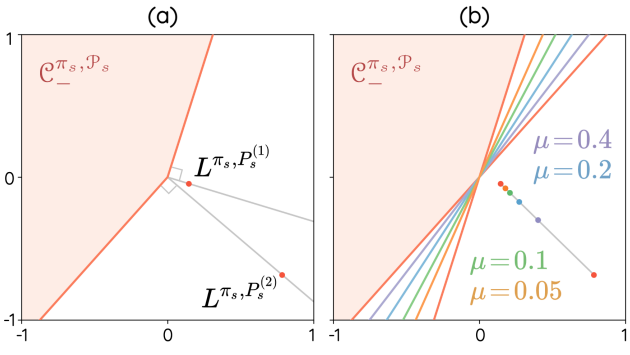

Figure 9. Visualization of the convex cone $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ for a fixed π_s and different $\mathcal{P}_s$. The translation $r^{\pi_s}\mathbf{1}$ is ignored since π_s is fixed. **(a)** We set $\mathcal{P}_s = \{P_s^{(1)}, P_s^{(2)}\}$. **(b)** We set $\mathcal{P}_s = \{P_s^{(\mu)} \mid P_s^{(\mu)} = \mu P_s^{(1)} + (1-\mu)P_s^{(2)}, 0 \le \mu \le 1\}$ and also plot the hyperplanes $H^{\pi_s,P_s^{(\mu)}}$ for different μ.

(see Appendix A). Our results also provide a geometric perspective on why the optimal policies under s-rectangularity might be stochastic, which is only exemplified in prior works (Wiesemann et al., 2013). The robust value functions of deterministic policies always lie in the region defined by RHS of Eqn. (34) but the optimal value might lie outside.

Theorem 4.8. *Consider an s-rectangular uncertainty set $\mathcal{P}$, the robust value function space $\mathcal{V}^{\mathcal{P}}$ satisfies*

$$\begin{aligned}\mathcal{V}^{\mathcal{P}} &= \bigcap_{s\in\mathcal{S}} \left[\left[\bigcup_{a\in\mathcal{A}} \mathcal{C}_+^{d_{s,a},\mathcal{P}_s}\right] \cap \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_-^{\pi_s,\mathcal{P}_s}\right]\right] \\ &\supseteq \bigcap_{s\in\mathcal{S}} \left[\left[\bigcup_{a\in\mathcal{A}} \mathcal{C}_+^{d_{s,a},\mathcal{P}_s}\right] \cap \left[\bigcup_{a\in\mathcal{A}} \mathcal{C}_-^{d_{s,a},\mathcal{P}_s}\right]\right]\end{aligned} \tag{34}$$

where the equality in the second line holds when $\mathcal{P}$ is (s,a)-rectangular.

Furthermore, we take a closer look at this "extra" region under s-rectangularity. Since the space can be decomposed state-wisely, we focus on a single state s. Recall the definition of $\mathcal{C}^{\pi_s,\mathcal{P}_s}$ in Eqn. (27), *i.e.*,

$$\mathcal{C}^{\pi_s,\mathcal{P}_s} = \bigcap_{P_s\in\mathcal{P}_s} \mathcal{H}_-^{\pi_s,P_s}. \tag{35}$$

From our results in Section 3, we know

$$\mathcal{H}_-^{\pi_s,P_s} \subseteq \bigcup_{a\in\mathcal{A}} \mathcal{H}_-^{d_{s,a},P_s}. \tag{36}$$

Therefore, we can obtain

$$\mathcal{C}^{\pi_s,\mathcal{P}_s} \subseteq \bigcap_{P_s\in\mathcal{P}_s} \bigcup_{a\in\mathcal{A}} \mathcal{H}_-^{d_{s,a},P_s}, \tag{37}$$

and accordingly

$$\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}^{\pi_s,\mathcal{P}_s} \subseteq \bigcap_{P_s\in\mathcal{P}_s} \bigcup_{a\in\mathcal{A}} \mathcal{H}_-^{d_{s,a},P_s}. \tag{38}$$

The RHS of the above equation gives us an upper bound of the region $\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}^{\pi_s,\mathcal{P}_s}$ while the RHS of Eqn. (33) provides a lower bound. The "extra" region lies within the gap between them. Figure 8 gives an illustration using the same RMDP example as in Figure 7.

4.3. Active Uncertainty Subsets

In above sections, we have shown that the robust value space $\mathcal{V}^{\mathcal{P}}$ depends on $\mathcal{P}$ in the form of a set of conic hypersurfaces $C^{\pi_s,\mathcal{P}_s}$. In this section, by taking a closer look at how $\mathcal{P}_s$ and $C^{\pi_s,\mathcal{P}_s}$ are related, we will show that only a subset $\mathcal{P}^\dagger \subseteq \mathcal{P}$ is sufficient to determine the robust value space, *i.e.*,

$$\mathcal{V}^{\mathcal{P}} = \mathcal{V}^{\mathcal{P}^\dagger}. \tag{39}$$

We term $\mathcal{P}^\dagger$ as *active* uncertainty subset, analogous to active constraints, in the sense that all $P \in \mathcal{P}^\dagger$ are active in determining the shape of the robust value space $\mathcal{V}^{\mathcal{P}}$.

First, let us keep π_s fixed, and note that the conic hypersurface $C^{\pi_s,\mathcal{P}_s}$ is uniquely determined by the convex cone $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$. We then focus on the relationship between $\mathcal{P}_s$ and $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$. Denote the set

$$\mathcal{L}^{\pi_s,\mathcal{P}_s} := \{L^{\pi_s,P_s} \mid P_s \in \mathcal{P}_s\}. \tag{40}$$

From the definition of $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$, we can see $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ is exactly the polar cone of $\mathcal{L}^{\pi_s,\mathcal{P}_s}$ (plus a translation), denoted with

$$\mathcal{C}_-^{\pi_s,\mathcal{P}_s} = (\mathcal{L}^{\pi_s,\mathcal{P}_s})^* + \{r^{\pi_s}\mathbf{1}\}. \tag{41}$$

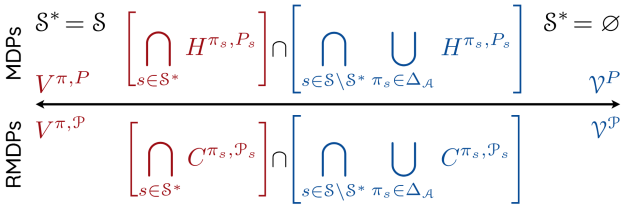

Figure 10. A spectrum of the spaces of value functions.

Here $+$ denotes the Minkowski addition. Figure 9(a) gives an illustration. Note that for fixed π_s, $\mathcal{L}^{\pi_s,\mathcal{P}_s}$ is the image of $\mathcal{P}_s$ under a fixed affine transformation. We denote this affine transformation as g, *i.e.*, $\mathcal{L}^{\pi_s,\mathcal{P}_s} = g(\mathcal{P}_s)$. Then we are able to obtain the following lemma:

Lemma 4.9. *Consider a s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have*

$$(\mathcal{L}^{\pi_s,\mathcal{P}_s})^* = (g(\mathcal{P}_s))^* = (g(\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s))))^*. \quad (42)$$

This lemma implies that, in order to determine the conic hypersurface $C^{\pi_s,\mathcal{P}_s}$, we only need to care about those $P_s \in \mathcal{P}_s$ that are extreme points of the convex hull. Figure 9(b) gives an illustration. We then generalize it to the whole robust value space and present the following theorem:

Theorem 4.10. *Consider a s-rectangular uncertainty set $\mathcal{P}$, we have*

$$\mathcal{V}^{\mathcal{P}} = \mathcal{V}^{\mathcal{P}^\dagger} \quad (43)$$

where $\mathcal{P}^\dagger = \mathbf{ext}(\mathbf{conv}(\mathcal{P})) \subseteq \mathcal{P}$.

If the $\mathcal{P}$ (or more generally $\mathbf{conv}(\mathcal{P})$) is polyhedral, such as ℓ_1-ball and ℓ_∞-ball (Ho et al., 2018; 2021; Behzadian et al., 2021), then we can reduce $\mathcal{P}$ to a finite set without losing any useful information for policy optimization. In addition, $\mathbf{conv}(\mathcal{P})$ being polyhedral implies that $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ is a polyhedral cone. Combining with Theorem 4.8, it means that the robust value space for an (s,a)-rectangular uncertainty set will be a polytope.

5. Discussion

5.1. Policy Agreement on More States

We already know that the value functions for policies that agree on a single state lie in a hyperplane for MDPs (Dadashi et al., 2019), and a conic hypersurface for s-rectangular RMDPs (Section 4.1). One natural question is how the space of value functions looks like when we fix the policies at more states. With our new decomposition-based perspective, the results are immediately available from Lemma 3.2 and Lemma 4.5.

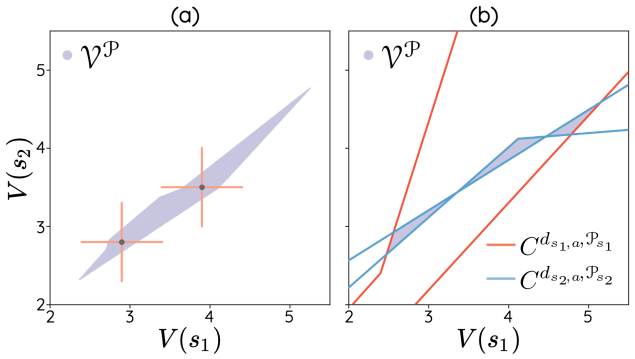

Figure 11. **(a)** The intersection between the robust value space and axis-parallel lines are line segments. **(b)** An example showing that the robust value space is not star-convex.

In Figure 10, we show the space of value functions for policies agree on states in $\mathcal{S}^* \subseteq \mathcal{S}$, under both non-robust and robust setting. Moreover, as illustrated in Figure 10, our perspective reveals a spectrum of the spaces of value functions. When the policies agree on all states, then it reduces to a single value function. When the policies are free to vary on all states, then it is the whole value space. This perspective enables us to characterize every point on this spectrum in an explicit form. In comparison, for non-robust case, prior works (Dadashi et al., 2019) only prove that the spaces are polytopes without giving a clear characterization.

5.2. The Non-convexity of the Robust Value Space

Like the non-robust case, the robust value space $\mathcal{V}^{\mathcal{P}}$ is also possibly non-convex (*e.g.*, Figure 7). Despite the non-convexity, $\mathcal{V}^{\mathcal{P}}$ exhibits some interesting properties analogous to monotone polygons. As shown in Figure 11(a), for any point in the robust value space $\mathcal{V}^{\mathcal{P}}$, if we draw an axis-parallel line passing this point, the intersection will be a line segment (or a point in degenerated case). We formalize this observation in the following corollary.

Corollary 5.1. *Consider an s-rectangular uncertainty set $\mathcal{P}$, if an axis-parallel line intersects with the robust value space $\mathcal{V}^{\mathcal{P}}$, then the intersection will be a line segment.*

From the examples in Figure 7, one may wonder if the robust value function space $\mathcal{V}^{\mathcal{P}}$ is a star-convex set. For many randomly generated RMDPs, $\mathcal{V}^{\mathcal{P}}$ does look like a star-convex set (see Figure 12 in Appendix C). However, we show a carefully crafted counter-example in Figure 11(b), which is clearly not star-shaped. Nevertheless, it seems to be a rare case. One interesting question to explore in the future is, how non-convex the robust value space can be and how likely it exhibits such non-convexity. If it is nearly convex for most time, then we might be able to design some efficient algorithms tailored for such case.

5.3. The Line Theorem for RMDPs

As mentioned before, one major obstacle that prevents us from adapting the prior method (Dadashi et al., 2019) from MDPs to RMDPs is that deriving a robust counterpart of the Line Theorem is highly challenging. Here we elaborate on this issue, with the help of our findings about the robust value space. Without loss of generality, suppose the set of policies only differ on s_1. From the discussions in Section 5.1, we know the resulting set of robust value functions is

$$\left[\bigcap_{i=2}^{S} C^{\pi_{s_i},\mathcal{P}_{s_i}}\right] \cap \left[\bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}} C^{\pi_{s_1},\mathcal{P}_{s_1}}\right]. \tag{44}$$

The first term is an intersection of $S-1$ conic hypersurfaces and the second term is an infinite union of conic hypersurfaces. Both are hard to further characterize. For example, though we know the first term could be a curve, it is challenging to give a closed-form expression for it. In comparison, for MDPs, the first term is just a line and its direction is known (see the proof of Lemma 4 (ii) in (Dadashi et al., 2019)).

6. Related Works

The geometry of the space of value functions has been studied only recently. Dadashi et al. (2019) first investigate it, and establish that for MDPs the value space is a possibly non-convex polytope. Their results provide a geometric perspective to help understand the dynamics of different RL algorithms (Kumar et al., 2019; Chan et al., 2020; Harb et al., 2020; Chan et al., 2021), and also inspire new methods in representation learning in RL (Bellemare et al., 2019; Dabney et al., 2021). In RMDP literature, some works take advantage of the geometric properties of special uncertainty sets to design efficient algorithms (Ho et al., 2018; Behzadian et al., 2021; Ho et al., 2021), but no prior works studies the geometry of the robust value space.

Our work can be viewed as an extension of (Dadashi et al., 2019) to RMDPs. We introduce a new perspective to characterize the geometric properties of the value space for RMDPs. Our approach also leads to a finer characterization of the value function polytope in MDPs setting.

7. Conclusion and Future Work

In this work, we characterize the geometry of the space of robust value functions from a new perspective, where the value space is decomposed in a state-wise manner. We show that the robust value space is determined by a set of conic hypersurfaces. Furthermore, we can reduce the uncertainty set to a subset of extreme points without sacrificing any useful information for policy optimization.

There remain some interesting open questions. As discussed in Section 5, it is worth studying how non-convex the robust value space can be (*i.e.*, can it be approximated as a convex set?). A further question is whether the level of non-convexity increases or decreases with the number of states/actions. Another direction is to investigate the geometry for other uncertainty set, such as coupled uncertainty (Mannor et al., 2012), r-rectangular sets (Goyal & Grand-Clément, 0) or more general ones. In addition, as in the non-robust case, it is interesting to study the geometry of robust value functions when the state space is very large and some approximation is needed. We will leave these questions to future works.

Acknowledgements

This work was partially supported by the Israel Science Foundation under contract 2199/20. We appreciate the valuable feedback from ICML anonymous reviewers. We also thank Bingyi Kang and Pengqian Yu for some helpful discussions about RMDPs.

References

Behzadian, B., Petrik, M., and Ho, C. P. Fast algorithms for l_∞-constrained s-rectangular robust mdps. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 25982–25992. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/file/da4fb5c6e93e74d3df8527599fa62642-Paper.pdf.

Bellemare, M., Dabney, W., Dadashi, R., Ali Taiga, A., Castro, P. S., Le Roux, N., Schuurmans, D., Lattimore, T., and Lyle, C. A geometric perspective on optimal representations for reinforcement learning. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/3cf2559725a9fdfa602ec8c887440f32-Paper.pdf.

Bellman, R. *Dynamic programming*. Princeton University Press, Princeton, 1957.

Bertsekas, D. *Convex Optimization Theory*. Athena Scientific optimization and computation series. Athena Scientific, 2009. ISBN 9781886529311. URL http://www.athenasc.com/convexduality.html.

Bertsekas, D., Nedic, A., and Ozdaglar, A. *Convex Analysis and Optimization*. Athena Scientific optimization

and computation series. Athena Scientific, 2003. ISBN 9781886529458. URL http://www.athenasc.com/convexity.html.

Chan, A., Asis, K. D., and Sutton, R. S. Inverse policy evaluation for value-based sequential decision-making. *CoRR*, abs/2008.11329, 2020. URL https://arxiv.org/abs/2008.11329.

Chan, A., Silva, H., Lim, S., Kozuno, T., Mahmood, A. R., and White, M. Greedification operators for policy optimization: Investigating forward and reverse KL divergences. *CoRR*, abs/2107.08285, 2021. URL https://arxiv.org/abs/2107.08285.

Dabney, W., Barreto, A., Rowland, M., Dadashi, R., Quan, J., G. Bellemare, M., and Silver, D. The value-improvement path: Towards better representations for reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8):7160–7168, May 2021. URL https://ojs.aaai.org/index.php/AAAI/article/view/16880.

Dadashi, R., Taiga, A. A., Roux, N. L., Schuurmans, D., and Bellemare, M. G. The value function polytope in reinforcement learning. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1486–1495. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/dadashi19a.html.

Dattorro, J. *Convex Optimization & Euclidean Distance Geometry*. Meboo Publishing, 2005. ISBN 9780976401308. URL https://meboo.convexoptimization.com/.

de Boer, P.-T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. A tutorial on the cross-entropy method. *Annals of Operations Research*, 134(1):19–67, Feb 2005. ISSN 1572-9338. doi: 10.1007/s10479-005-5724-z. URL https://doi.org/10.1007/s10479-005-5724-z.

Goyal, V. and Grand-Clément, J. Robust markov decision processes: Beyond rectangularity. *Mathematics of Operations Research*, 0(0):null, 0. doi: 10.1287/moor.2022.1259. URL https://doi.org/10.1287/moor.2022.1259.

Harb, J., Schaul, T., Precup, D., and Bacon, P. Policy evaluation networks. *CoRR*, abs/2002.11833, 2020. URL https://arxiv.org/abs/2002.11833.

Ho, C. P., Petrik, M., and Wiesemann, W. Fast Bellman updates for robust MDPs. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1979–1988. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/ho18a.html.

Ho, C. P., Petrik, M., and Wiesemann, W. Partial policy iteration for l1-robust markov decision processes. *Journal of Machine Learning Research*, 22(275):1–46, 2021. URL http://jmlr.org/papers/v22/20-445.html.

Howard, R. A. *Dynamic programming and Markov processes*. Dynamic programming and Markov processes. John Wiley, Oxford, England, 1960.

Iyengar, G. N. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/25151652.

Krein, M. and Milman, D. On extreme points of regular convex sets. *Studia Mathematica*, 9(1):133–138, 1940. URL http://eudml.org/doc/219061.

Kumar, S., Ahmed, Z., Dadashi, R., Schuurmans, D., and Bellemare, M. G. Generalized policy updates for policy optimization. In *NeurIPS 2019 Optimization Foundations for Reinforcement Learning Workshop*, 2019.

Mannor, S., Mebel, O., and Xu, H. Lightning does not strike twice: Robust mdps with coupled uncertainty. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pp. 451–458, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.

Nilim, A. and El Ghaoui, L. Robust control of markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005. doi: 10.1287/opre.1050.0216. URL https://doi.org/10.1287/opre.1050.0216.

Sutton, R. S., McAllester, D., Singh, S., and Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In Solla, S., Leen, T., and Müller, K. (eds.), *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 1999. URL https://proceedings.neurips.cc/paper/1999/file/464d828b85b0bed98e80ade0a5c43b0f-Paper.pdf.

Wiesemann, W., Kuhn, D., and Rustem, B. Robust markov decision processes. *Mathematics of Operations Research*, 38(1):153–183, 2013. doi: 10.1287/moor.1120.0566. URL https://doi.org/10.1287/moor.1120.0566.

Xu, H. and Mannor, S. The robustness-performance tradeoff in markov decision processes. In Schölkopf, B., Platt, J., and Hoffman, T. (eds.), *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2006. URL https://proceedings.neurips.cc/paper/2006/file/177540c7bcb8db31697b601642eac8d4-Paper.pdf.

A. Details of MDPs and RMDPs

In this section, we give the specifics of the MDPs and the RMDPs used for illustrations in this work.

Figure 2(a) and Figure 3:

$$
\begin{aligned}
&S = 2, A = 3 \\
&r_{s_1} = (0.0199, 0.6097, 0.8313), r_{s_2} = (0.4044, 0.5534, 0.8319) \\
&P_{s_1,a_1} = (0.7793, 0.2207), P_{s_1,a_2} = (0.9713, 0.0287), P_{s_1,a_3} = (0.0668, 0.9332) \\
&P_{s_2,a_1} = (0.0676, 0.9324), P_{s_2,a_2} = (0.5929, 0.4071), P_{s_2,a_3} = (0.2497, 0.7503) \\
&\pi_{s_1} = (0.2, 0.3, 0.5), \pi_{s_2} = (0.3, 0.1, 0.6)
\end{aligned}
$$

Figure 2(b):

$$
\begin{aligned}
&S = 3, A = 2 \\
&r_{s_1} = (0.5, 0.8), r_{s_2} = (0.4, 0.2), r_{s_3} = (0.2, 0.6) \\
&P_{s_1,a_1} = (0.14, 0.75, 0.11), P_{s_1,a_2} = (0.44, 0.45, 0.11) \\
&P_{s_2,a_1} = (0.23, 0.19, 0.58), P_{s_2,a_2} = (0.44, 0.32, 0.24) \\
&P_{s_3,a_1} = (0.45, 0.43, 0.12), P_{s_3,a_2} = (0.14, 0.54, 0.32) \\
&\pi_{s_1} = (0.46, 0.54), \pi_{s_2} = (0.38, 0.62), \pi_{s_3} = (0.49, 0.51)
\end{aligned}
$$

Figure 4:

$$
\begin{aligned}
&S = 2, A = 2 \\
&r_{s_1} = (0.5, 0.6), r_{s_2} = (0.4, 0.7) \\
&\mathcal{P}_{s_1} = \left\{ \begin{pmatrix} 0.78 & 0.22 \\ 0.79 & 0.21 \end{pmatrix}, \begin{pmatrix} 0.85 & 0.15 \\ 0.99 & 0.01 \end{pmatrix} \right\} \\
&\mathcal{P}_{s_2} = \left\{ \begin{pmatrix} 0.59 & 0.41 \\ 0.92 & 0.08 \end{pmatrix}, \begin{pmatrix} 0.60 & 0.40 \\ 0.39 & 0.61 \end{pmatrix} \right\} \\
&\pi_{s_1} = (0.45, 0.55), \pi_{s_2} = (0.10, 0.90)
\end{aligned}
$$

Figure 5:

$$
\begin{aligned}
&S = 2, A = 2 \\
&r_{s_1} = (0.5, 0.6), r_{s_2} = (0.4, 0.7) \\
&\mathcal{P}_{s_1} = \left\{ \begin{pmatrix} 0.78 & 0.22 \\ 0.79 & 0.21 \end{pmatrix}, \begin{pmatrix} 0.85 & 0.15 \\ 0.99 & 0.01 \end{pmatrix}, \begin{pmatrix} 0.92 & 0.08 \\ 0.99 & 0.01 \end{pmatrix}, \begin{pmatrix} 0.92 & 0.08 \\ 0.83 & 0.17 \end{pmatrix} \right\} \\
&\mathcal{P}_{s_2} = \left\{ \begin{pmatrix} 0.59 & 0.41 \\ 0.92 & 0.08 \end{pmatrix} \right\} \\
&\pi_{s_1} = (0.45, 0.55), \pi_{s_2} = (0.10, 0.90)
\end{aligned}
$$

Figure 6, Figure 7(a), Figure 8, Figure 9 and Figure 11(a):

$$
\begin{aligned}
&S = 2, A = 2 \\
&r_{s_1} = (0.27, 0.9398), r_{s_2} = (0.3374, 0.2212) \\
&\mathcal{P}_{s_1} = \left\{ \begin{pmatrix} 0.95 & 0.05 \\ 0.17 & 0.83 \end{pmatrix}, \begin{pmatrix} 0.24 & 0.76 \\ 0.05 & 0.95 \end{pmatrix} \right\} \\
&\mathcal{P}_{s_2} = \left\{ \begin{pmatrix} 0.07 & 0.93 \\ 0.83 & 0.17 \end{pmatrix}, \begin{pmatrix} 0.70 & 0.30 \\ 0.23 & 0.77 \end{pmatrix} \right\} \\
&\pi_{s_1} = (0.8, 0.2), \pi_{s_2} = (0.9, 0.1)
\end{aligned}
$$

Figure 7(b):

$$
\begin{aligned}
&S = 2, A = 2 \\
&r_{s_1} = (0.27, 0.9398), r_{s_2} = (0.3374, 0.2212) \\
&\mathcal{P}_{s_1,a_1} = \{(0.95, 0.05), (0.24, 0.76)\} \\
&\mathcal{P}_{s_1,a_2} = \{(0.17, 0.83), (0.05, 0.95)\} \\
&\mathcal{P}_{s_2,a_1} = \{(0.07, 0.93), (0.70, 0.30)\} \\
&\mathcal{P}_{s_1,a_1} = \{(0.83, 0.17), (0.23, 0.77)\}
\end{aligned}
$$

Figure 11(b):

$$
\begin{aligned}
&S = 2, A = 2 \\
&r_{s_1} = (0.24, 0.998), r_{s_2} = (0.3574, 0.412) \\
&\mathcal{P}_{s_1} = \left\{ \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}, \begin{pmatrix} 0.24 & 0.76 \\ 0.95 & 0.05 \end{pmatrix} \right\} \\
&\mathcal{P}_{s_2} = \left\{ \begin{pmatrix} 0.2 & 0.8 \\ 0.99 & 0.01 \end{pmatrix}, \begin{pmatrix} 0.2 & 0.8 \\ 0.01 & 0.99 \end{pmatrix} \right\}
\end{aligned}
$$

B. Proofs

Lemma 3.1. *Consider a policy π and a transition dynamic P, we have*

$$\{V^{\pi,P}\} = \bigcap_{s \in \mathcal{S}} H^{\pi_s, P_s} \tag{17}$$

Proof. Observe that

$$H^{\pi_s,P_s} = \{\mathbf{x} \in \mathbb{R}^{\mathcal{S}} \mid \langle \mathbf{x}, L^{\pi_s,P_s} \rangle = r^{\pi_s}\} \tag{45}$$

is the set of vectors that satisfy the s-th equation of the following system of linear equations:

$$(I - \gamma P^{\pi})\mathbf{x} = r^{\pi}. \tag{46}$$

Since $(I - \gamma P^{\pi})$ is invertible, this system of linear equations has a unique solution $V^{\pi,P}$. Hence, we have

$$\{V^{\pi,P}\} = \bigcap_{s \in \mathcal{S}} H^{\pi_s, P_s} \tag{47}$$

which completes the proof. □

Lemma 3.2. *Consider a transition dynamic P, the value space $\mathcal{V}^P$ can be represented as*

$$\mathcal{V}^P = \bigcup_{\pi \in \Pi} \bigcap_{s \in \mathcal{S}} H^{\pi_s, P_s} = \bigcap_{s \in \mathcal{S}} \bigcup_{\pi_s \in \Delta_{\mathcal{A}}} H^{\pi_s, P_s}. \tag{19}$$

Proof. By the definition of $\mathcal{V}^P$ and Lemma 3.1, we have

$$\mathcal{V}^P = \bigcup_{\pi \in \Pi} \{V^{\pi,P}\} = \bigcup_{\pi \in \Pi} \bigcap_{s \in \mathcal{S}} H^{\pi_s, P_s}. \tag{48}$$

We can break the union into nested unions by fixing π_s for each s:

$$\bigcup_{\pi \in \Pi} \bigcap_{s \in \mathcal{S}} H^{\pi_s, P_s} = \bigcup_{\pi_{s_S} \in \Delta_{\mathcal{A}}} \cdots \bigcup_{\pi_{s_2} \in \Delta_{\mathcal{A}}} \bigcup_{\pi_{s_1} \in \Delta_{\mathcal{A}}} \bigcap_{s \in \mathcal{S}} H^{\pi_s, P_s}. \tag{49}$$

Then, we have

$$\begin{aligned}\mathcal{V}^P &= \bigcup_{\pi_{s_S}\in\Delta_{\mathcal{A}}} \cdots \bigcup_{\pi_{s_2}\in\Delta_{\mathcal{A}}} \bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}} \bigcap_{s\in\mathcal{S}} H^{\pi_s,P_s} \\ &= \bigcup_{\pi_{s_S}\in\Delta_{\mathcal{A}}} \cdots \bigcup_{\pi_{s_2}\in\Delta_{\mathcal{A}}} \bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}} \left[H^{\pi_{s_1},P_{s_1}} \cap \left[\bigcap_{i=2}^{S} H^{\pi_{s_i},P_{s_i}} \right] \right] \\ &= \bigcup_{\pi_{s_S}\in\Delta_{\mathcal{A}}} \cdots \bigcup_{\pi_{s_2}\in\Delta_{\mathcal{A}}} \left[\left[\bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}} H^{\pi_{s_1},P_{s_1}} \right] \cap \left[\bigcap_{i=2}^{S} H^{\pi_{s_i},P_{s_i}} \right] \right]. \quad \text{(distributive law of sets)}\end{aligned} \tag{50}$$

By iteratively applying the distributive law of sets, we can obtain

$$\mathcal{V}^P = \bigcap_{s\in\mathcal{S}} \bigcup_{\pi_s\in\Delta_{\mathcal{A}}} H^{\pi_s,P_s} \tag{51}$$

which completes the proof. □

Lemma 3.3. *Consider a policy π and a transition dynamic P, we have for all states $s \in \mathcal{S}$,*

$$\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{H}_+^{\pi_s,P_s} \cap \mathcal{H}_-^{\pi_s,P_s} = \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{H}_+^{\pi_s,P_s} \right] \cap \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{H}_-^{\pi_s,P_s} \right]. \tag{22}$$

Proof. First, by the distributive property of sets, it is trivial to obtain LHS $\subseteq$ RHS. Next, we will show RHS $\subseteq$ LHS. For any $\mathbf{x} \in$ RHS, we have

$$\exists\, \pi_s', \pi_s'' \in \Delta_{\mathcal{A}} \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{H}_+^{\pi_s',P_s} \cap \mathcal{H}_-^{\pi_s'',P_s}. \tag{52}$$

When $\pi_s' = \pi_s''$, it is trivial to obtain $\mathbf{x} \in$ LHS. When $\pi_s' \neq \pi_s''$, then there exists $\alpha, \beta \geq 0$ such that

$$\begin{aligned}\langle \mathbf{x}, L^{\pi_s',P_s} \rangle - r^{\pi_s'} &= \alpha, \\ \langle \mathbf{x}, L^{\pi_s'',P_s} \rangle - r^{\pi_s''} &= -\beta.\end{aligned} \tag{53}$$

When either $\alpha = 0$ or $\beta = 0$, we have $\mathbf{x} \in H^{\pi_s',P_s}$ or $\mathbf{x} \in H^{\pi_s'',P_s}$, and accordingly $\mathbf{x} \in$ LHS. Therefore, we only focus on the case where $\alpha, \beta > 0$. If we set

$$\pi_s^\dagger = \frac{\beta}{\alpha+\beta}\pi_s' + \frac{\alpha}{\alpha+\beta}\pi_s'', \tag{54}$$

then we have

$$\begin{aligned}&\langle \mathbf{x}, L^{\pi_s^\dagger,P_s} \rangle - r^{\pi_s^\dagger} \\ =& \langle \mathbf{x}, \frac{\beta}{\alpha+\beta} L^{\pi_s',P_s} + \frac{\alpha}{\alpha+\beta} L^{\pi_s'',P_s} \rangle - \frac{\beta}{\alpha+\beta} r^{\pi_s'} - \frac{\alpha}{\alpha+\beta} r^{\pi_s''} \\ =& \langle \mathbf{x}, \frac{\beta}{\alpha+\beta} L^{\pi_s',P_s} \rangle - \frac{\beta}{\alpha+\beta} r^{\pi_s'} + \langle \mathbf{x}, \frac{\alpha}{\alpha+\beta} L^{\pi_s'',P_s} \rangle - \frac{\alpha}{\alpha+\beta} r^{\pi_s''} \\ =& \frac{\beta}{\alpha+\beta} \left(\langle \mathbf{x}, L^{\pi_s',P_s} \rangle - r^{\pi_s'} \right) + \frac{\alpha}{\alpha+\beta} \left(\langle \mathbf{x}, L^{\pi_s'',P_s} \rangle - r^{\pi_s''} \right) \\ =& 0.\end{aligned} \tag{55}$$

Note that $\pi_s^\dagger \in \Delta_{\mathcal{A}}$. The above result implies $\mathbf{x}$ lies in the hyperplane $H^{\pi_s^\dagger,P_s}$. Thus $\mathbf{x} \in$ LHS and accordingly RHS $\subseteq$ LHS. Putting it together, we obtain LHS = RHS. □

Lemma 3.4. *Consider a policy π and a transition dynamic P, we have for all states $s \in \mathcal{S}$,*

$$\bigcup_{\pi_s \in \Delta_{\mathcal{A}}} \mathcal{H}_\delta^{\pi_s, P_s} = \bigcup_{a \in \mathcal{A}} \mathcal{H}_\delta^{d_{s,a}, P_s}, \quad \forall\, \delta \in \{+, -\}. \tag{23}$$

Proof. We first prove $\bigcup_{a\in\mathcal{A}} \mathcal{H}_+^{d_{s,a},P_s} = \bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{H}_+^{\pi_s,P_s}$. It is trivial that LHS $\subseteq$ RHS. We then focus on proving RHS $\subseteq$ LHS. For any $\mathbf{x} \in$ RHS, we have

$$\exists\, \pi'_s \in \Delta_{\mathcal{A}} \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{H}_+^{\pi'_s, P_s}. \tag{56}$$

Note that any $\pi_s \in \Delta_{\mathcal{A}}$ can be written as a convex combination of $d_{s,a}, a \in \mathcal{A}$. In our case, we write

$$\pi'_s = \sum_{a\in\mathcal{A}} \pi'_{s,a} d_{s,a}, \tag{57}$$

then we have

$$\begin{aligned} \langle \mathbf{x}, L^{\pi'_s, P_s} \rangle - r^{\pi'_s} &\geq 0 \\ \langle \mathbf{x}, \sum_{a\in\mathcal{A}} \pi'_{s,a} L^{d_{s,a},P_s} \rangle - \sum_{a\in\mathcal{A}} \pi'_{s,a} r^{d_{s,a}} &\geq 0 \\ \sum_{a\in\mathcal{A}} \pi'_{s,a} \left(\langle \mathbf{x}, L^{d_{s,a},P_s} \rangle - r^{d_{s,a}} \right) &\geq 0. \end{aligned} \tag{58}$$

Since $\pi'_{s,a} \geq 0$ for all $a \in \mathcal{A}$, the above inequality implies

$$\exists\, a' \in \mathcal{A} \quad \text{s.t.} \quad \langle \mathbf{x}, L^{d_{s,a'},P_s} \rangle - r^{d_{s,a'}} \geq 0. \tag{59}$$

This is equivalent to $\mathbf{x} \in$ LHS. Putting it together, we obtain LHS $=$ RHS.

The second part $\bigcup_{a\in\mathcal{A}} \mathcal{H}_-^{d_{s,a},P_s} = \bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{H}_-^{\pi_s,P_s}$ can be proved in the same way. □

Theorem 3.5. *Consider a transition dynamic P, the value space $\mathcal{V}^P$ can be represented as*

$$\begin{aligned} \mathcal{V}^P &= \bigcap_{s\in\mathcal{S}} \left[\left[\bigcup_{a\in\mathcal{A}} \mathcal{H}_+^{d_{s,a},P_s} \right] \cap \left[\bigcup_{a\in\mathcal{A}} \mathcal{H}_-^{d_{s,a},P_s} \right] \right] \\ &= \bigcup_{\mathbf{a}\in\mathcal{A}^{\mathcal{S}}} \bigcup_{\mathbf{a}'\in\mathcal{A}^{\mathcal{S}}} \bigcap_{s\in\mathcal{S}} \left[\mathcal{H}_+^{d_{s,\mathbf{a}_s},P_s} \cap \mathcal{H}_-^{d_{s,\mathbf{a}'_s},P_s} \right] \end{aligned} \tag{24}$$

where $\mathbf{a} = (\mathbf{a}_s)_{s\in\mathcal{S}}$, $\mathbf{a}' = (\mathbf{a}'_s)_{s\in\mathcal{S}}$, and $\mathbf{a}_s, \mathbf{a}'_s \in \mathcal{A}$.

Proof. The first equality follow immediately from Lemma 3.2, Lemma 3.3 and Lemma 3.4. The second equality can be obtained using the distributive law of sets. □

Lemma 4.1. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have for all states $s \in \mathcal{S}$,*

$$f_{\mathcal{P}}(Y^{\pi_s}) \subseteq \left[\bigcap_{P_s\in\mathcal{P}_s} \mathcal{H}_-^{\pi_s,P_s} \right] \cap \left[\bigcup_{P_s\in\mathcal{P}_s} H^{\pi_s,P_s} \right]. \tag{25}$$

Proof. For any $\pi' \in Y^{\pi_s}$, from Eqn. (9), we know that

$$\exists\, P_\dagger \in \mathcal{P}, \quad \text{s.t.} \quad \forall P \in \mathcal{P},\, V^{\pi',P_\dagger} \leq V^{\pi',P}. \tag{60}$$

Using the Bellman equation (Bellman, 1957), we can obtain

$$\begin{aligned} V^{\pi',P_\dagger} - V^{\pi',P} &= \gamma P_\dagger^{\pi'} V^{\pi',P_\dagger} - \gamma P^{\pi'} V^{\pi',P} \\ &= \gamma (P_\dagger^{\pi'} - P^{\pi'}) V^{\pi',P_\dagger} - \gamma P^{\pi'} (V^{\pi',P_\dagger} - V^{\pi',P}) \\ &= (I - \gamma P^{\pi'})^{-1} \gamma (P_\dagger^{\pi'} - P^{\pi'}) V^{\pi',P_\dagger} \end{aligned} \tag{61}$$

Note that $(I - \gamma P^{\pi'})^{-1} = \sum_{t=0}^{\infty} (\gamma P^{\pi'})^t \geq 0$. Thus, we have

$$\forall P \in \mathcal{P}, \quad \gamma(P_\dagger^{\pi'} - P^{\pi'}) V^{\pi', P_\dagger} \leq 0 \tag{62}$$

Rearranging the above inequality, we obtain

$$\forall P \in \mathcal{P}, \quad (I - \gamma P^{\pi'}) V^{\pi', P_\dagger} \leq (I - \gamma P_\dagger^{\pi'}) V^{\pi', P_\dagger}. \tag{63}$$

Since $(I - \gamma P_\dagger^{\pi'}) V^{\pi', P_\dagger} = r^{\pi'}$ and $V^{\pi', \mathcal{P}} = V^{\pi', P_\dagger}$, we have

$$\forall P \in \mathcal{P}, \quad (I - \gamma P^{\pi'}) V^{\pi', \mathcal{P}} \leq r^{\pi'}. \tag{64}$$

Taking the s-th inequality and noting that $\pi'_s = \pi_s$, we have

$$\forall P_s \in \mathcal{P}_s, \quad \langle V^{\pi', \mathcal{P}}, L^{\pi_s, P_s} \rangle \leq r^{\pi_s}. \tag{65}$$

Therefore, we have

$$f_{\mathcal{P}}(Y^{\pi_s}) \subseteq \bigcap_{P_s \in \mathcal{P}_s} \mathcal{H}_-^{\pi_s, P_s}. \tag{66}$$

On the other hand, from Eqn. (9) we know

$$\exists\, P_s \in \mathcal{P}_s, \quad \langle V^{\pi', \mathcal{P}}, L^{\pi_s, P_s} \rangle = r^{\pi_s}, \tag{67}$$

which is equivalent to

$$f_{\mathcal{P}}(Y^{\pi_s}) \subseteq \bigcup_{P_s \in \mathcal{P}_s} H^{\pi_s, P_s}. \tag{68}$$

Putting it together, we get

$$f_{\mathcal{P}}(Y^{\pi_s}) \subseteq \left[\bigcap_{P_s \in \mathcal{P}_s} \mathcal{H}_-^{\pi_s, P_s} \right] \cap \left[\bigcup_{P_s \in \mathcal{P}_s} H^{\pi_s, P_s} \right], \tag{69}$$

which completes the proof. □

Lemma 4.2. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have for all states $s \in \mathcal{S}$,*

$$\frac{r^{\pi_s}}{1-\gamma} \mathbf{1} \in \bigcap_{P_s \in \mathcal{P}_s} H^{\pi_s, P_s}. \tag{26}$$

Proof. Recall that

$$H^{\pi_s, P_s} = \{ \mathbf{x} \in \mathbb{R}^{\mathcal{S}} \mid \langle \mathbf{x}, L^{\pi_s, P_s} \rangle = r^{\pi_s} \}. \tag{70}$$

From the definition of L^{π_s, P_s}, we know

$$\langle \mathbf{1}, L^{\pi_s, P_s} \rangle = \frac{1}{1-\gamma}. \tag{71}$$

Thus, it is easy to verify that $\langle \frac{r^{\pi_s}}{1-\gamma} \mathbf{1}, L^{\pi_s, P_s} \rangle = r^{\pi_s}$ for all $P_s \in \mathcal{P}_s$, which concludes the proof. □

Corollary 4.3. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have for all states $s \in \mathcal{S}$,*

$$f_{\mathcal{P}}(Y^{\pi_s}) \subseteq C^{\pi_s, \mathcal{P}_s} \tag{28}$$

where $C^{\pi_s, \mathcal{P}_s} = \mathcal{C}_+^{\pi_s, \mathcal{P}_s} \cap \mathcal{C}_-^{\pi_s, \mathcal{P}_s}$ is a conic hypersurface.

Proof. This corollary is a restatement of Lemma 4.1. Note that

$$\begin{aligned}\left[\bigcap_{P_s\in\mathcal{P}_s}\mathcal{H}_-^{\pi_s,P_s}\right]\cap\left[\bigcup_{P_s\in\mathcal{P}_s}H^{\pi_s,P_s}\right]&=\bigcup_{P_s\in\mathcal{P}_s}\left[\left[\bigcap_{P_s\in\mathcal{P}_s}\mathcal{H}_-^{\pi_s,P_s}\right]\cap H^{\pi_s,P_s}\right]\\&=\bigcup_{P_s\in\mathcal{P}_s}\left[\left[\bigcap_{P_s\in\mathcal{P}_s}\mathcal{H}_-^{\pi_s,P_s}\right]\cap\mathcal{H}_+^{\pi_s,P_s}\right]\\&=\left[\bigcap_{P_s\in\mathcal{P}_s}\mathcal{H}_-^{\pi_s,P_s}\right]\cap\left[\bigcup_{P_s\in\mathcal{P}_s}\mathcal{H}_+^{\pi_s,P_s}\right]\\&=\mathcal{C}_-^{\pi_s,\mathcal{P}_s}\cap\mathcal{C}_+^{\pi_s,\mathcal{P}_s}.\end{aligned}\tag{72}$$

From Lemma 4.2, we know all halfspaces $\mathcal{H}_-^{\pi_s,P_s}$ intersect at the same point. Then their intersection $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ will be a convex cone. Note that each H^{π_s,P_s} is a supporting hyperplane of the cone $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ and all H^{π_s,P_s} determine this cone. Thus the intersection of $\bigcup_{P_s\in\mathcal{P}_s}H^{\pi_s,P_s}$ and $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$ is exactly the surface of $\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$. □

Lemma 4.4. *Consider an s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have*

$$\{V^{\pi,\mathcal{P}}\}=\bigcap_{s\in\mathcal{S}}C^{\pi_s,\mathcal{P}_s}.\tag{29}$$

Proof. For any $\mathbf{x}\in$ RHS, we have that for all $s\in\mathcal{S}$

$$\exists\, P_s\in\mathcal{P}_s,\quad \langle\mathbf{x},L^{\pi_s,P_s}\rangle=r^{\pi_s};\tag{73}$$
$$\forall\, P_s\in\mathcal{P}_s,\quad \langle\mathbf{x},L^{\pi_s,P_s}\rangle\le r^{\pi_s}.\tag{74}$$

Since P is s-rectangular, we have

$$\exists\, P\in\mathcal{P},\quad (I-\gamma P^\pi)\mathbf{x}=r^\pi;\tag{75}$$
$$\forall\, P\in\mathcal{P},\quad (I-\gamma P^\pi)\mathbf{x}\le r^\pi.\tag{76}$$

Since the Bellman equation has a unique solution, the first line implies $\exists\, P_\dagger\in\mathcal{P},\mathbf{x}=V^{\pi,P_\dagger}$. Suppose $V^{\pi,P_\dagger}\neq V^{\pi,\mathcal{P}}$, then from Eqn. (9) we have

$$\exists\, P_\ddagger\in\mathcal{P},\quad \text{s.t.}\quad V^{\pi,P_\ddagger}=V^{\pi,\mathcal{P}}<V^{\pi,P_\dagger}.\tag{77}$$

On the other hand, from Eqn. (76), we know

$$\begin{aligned}(I-\gamma P_\ddagger^\pi)V^{\pi,P_\dagger}-r^\pi&\le 0\\(I-\gamma P_\ddagger^\pi)V^{\pi,P_\dagger}-(I-\gamma P_\dagger^\pi)V^{\pi,P_\dagger}&\le 0\\\gamma(P_\dagger^\pi-P_\ddagger^\pi)V^{\pi,P_\dagger}&\le 0\\(I-\gamma P_\ddagger^\pi)^{-1}\gamma(P_\dagger^\pi-P_\ddagger^\pi)V^{\pi,P_\dagger}&\le 0\qquad\text{(see the proof of Lemma 4.1)}\\V^{\pi,P_\dagger}-V^{\pi,P_\ddagger}&\le 0\\V^{\pi,P_\dagger}&\le V^{\pi,P_\ddagger}.\end{aligned}\tag{78}$$

We have an contradiction. Therefore, we can conclude $\mathbf{x}=V^{\pi,\mathcal{P}}$ and accordingly $\{V^{\pi,\mathcal{P}}\}=\bigcap_{s\in\mathcal{S}}C^{\pi_s,\mathcal{P}_s}$. □

Lemma 4.5. *Consider an s-rectangular uncertainty set $\mathcal{P}$, the robust value function space $\mathcal{V}^{\mathcal{P}}$ can be represented as*

$$\mathcal{V}^{\mathcal{P}}=\bigcup_{\pi\in\Pi}\bigcap_{s\in\mathcal{S}}C^{\pi_s,\mathcal{P}_s}=\bigcap_{s\in\mathcal{S}}\bigcup_{\pi_s\in\Delta_{\mathcal{A}}}C^{\pi_s,\mathcal{P}_s}.\tag{30}$$

Proof. The proof below follows exactly the same procedure as the proof of Lemma 3.2. By the definition of $\mathcal{V}^{\mathcal{P}}$ and Lemma 4.4, we have

$$\mathcal{V}^{\mathcal{P}} = \bigcup_{\pi\in\Pi}\{V^{\pi,\mathcal{P}}\} = \bigcup_{\pi\in\Pi}\bigcap_{s\in\mathcal{S}} C^{\pi_s,\mathcal{P}_s}. \tag{79}$$

We can break the union into nested unions by fixing π_s for each s:

$$\bigcup_{\pi\in\Pi}\bigcap_{s\in\mathcal{S}} C^{\pi_s,\mathcal{P}_s} = \bigcup_{\pi_{s_S}\in\Delta_{\mathcal{A}}}\cdots\bigcup_{\pi_{s_2}\in\Delta_{\mathcal{A}}}\bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}}\bigcap_{s\in\mathcal{S}} C^{\pi_s,\mathcal{P}_s}. \tag{80}$$

Then, we have

$$\begin{aligned}
\mathcal{V}^{\mathcal{P}} &= \bigcup_{\pi_{s_S}\in\Delta_{\mathcal{A}}}\cdots\bigcup_{\pi_{s_2}\in\Delta_{\mathcal{A}}}\bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}}\bigcap_{s\in\mathcal{S}} C^{\pi_s,\mathcal{P}_s} \\
&= \bigcup_{\pi_{s_S}\in\Delta_{\mathcal{A}}}\cdots\bigcup_{\pi_{s_2}\in\Delta_{\mathcal{A}}}\bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}}\left[C^{\pi_{s_1},\mathcal{P}_{s_1}}\cap\left[\bigcap_{i=2}^{S} C^{\pi_{s_i},\mathcal{P}_{s_i}}\right]\right] \\
&= \bigcup_{\pi_{s_S}\in\Delta_{\mathcal{A}}}\cdots\bigcup_{\pi_{s_2}\in\Delta_{\mathcal{A}}}\left[\left[\bigcup_{\pi_{s_1}\in\Delta_{\mathcal{A}}} C^{\pi_{s_1},\mathcal{P}_{s_1}}\right]\cap\left[\bigcap_{i=2}^{S} C^{\pi_{s_i},\mathcal{P}_{s_i}}\right]\right]. \quad \text{(distributive law of sets)}
\end{aligned} \tag{81}$$

By iteratively applying the distributive law of sets, we can obtain

$$\mathcal{V}^{\mathcal{P}} = \bigcap_{s\in\mathcal{S}}\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} C^{\pi_s,\mathcal{P}_s}, \tag{82}$$

which completes the proof. □

Lemma 4.6. *Consider an s-rectangular uncertainty set $\mathcal{P}$, we have for all states $s\in\mathcal{S}$,*

$$\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} C^{\pi_s,\mathcal{P}_s} = \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}}\mathcal{C}_+^{\pi_s,\mathcal{P}_s}\right]\cap\left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}}\mathcal{C}_-^{\pi_s,\mathcal{P}_s}\right]. \tag{31}$$

Proof. Recall that $C^{\pi_s,\mathcal{P}_s} = \mathcal{C}_+^{\pi_s,\mathcal{P}_s}\cap\mathcal{C}_-^{\pi_s,\mathcal{P}_s}$, then we need to prove

$$\bigcup_{\pi_s\in\Delta_{\mathcal{A}}}\mathcal{C}_+^{\pi_s,\mathcal{P}_s}\cap\mathcal{C}_-^{\pi_s,\mathcal{P}_s} = \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}}\mathcal{C}_+^{\pi_s,\mathcal{P}_s}\right]\cap\left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}}\mathcal{C}_-^{\pi_s,\mathcal{P}_s}\right]. \tag{83}$$

First, by the distributive property of sets, it is trivial to obtain LHS $\subseteq$ RHS. Next, we will show RHS $\subseteq$ LHS. For any $\mathbf{x}\in$ RHS, we have

$$\exists\,\pi_s',\pi_s''\in\Delta_{\mathcal{A}} \quad \text{s.t.} \quad \mathbf{x}\in\mathcal{C}_+^{\pi_s',\mathcal{P}_s}\cap\mathcal{C}_-^{\pi_s'',\mathcal{P}_s}. \tag{84}$$

When $\pi_s' = \pi_s''$, it is trivial to obtain $\mathbf{x}\in$ LHS. When $\pi_s'\neq\pi_s''$, then we have

$$\begin{aligned}
&\exists\, P_s\in\mathcal{P}_s, \quad \langle\mathbf{x}, L^{\pi_s',P_s}\rangle - r^{\pi_s'}\geq 0; \\
&\forall\, P_s\in\mathcal{P}_s, \quad \langle\mathbf{x}, L^{\pi_s'',P_s}\rangle - r^{\pi_s''}\leq 0.
\end{aligned} \tag{85}$$

If there exists $P_s\in\mathcal{P}_s$ such that $\langle\mathbf{x}, L^{\pi_s'',P_s}\rangle - r^{\pi_s''} = 0$, then we will get $\mathbf{x}\in H^{\pi_s'',P_s}\subseteq\mathcal{C}_+^{\pi_s'',\mathcal{P}_s}$ and accordingly $\mathbf{x}\in$ LHS. Therefore, we only consider the case where

$$\begin{aligned}
&\exists\, P_s\in\mathcal{P}_s, \quad \langle\mathbf{x}, L^{\pi_s',P_s}\rangle - r^{\pi_s'}\geq 0; \\
&\forall\, P_s\in\mathcal{P}_s, \quad \langle\mathbf{x}, L^{\pi_s'',P_s}\rangle - r^{\pi_s''}< 0.
\end{aligned} \tag{86}$$

We denote

$$\begin{aligned}
\alpha^{P_s} &\coloneqq \langle \mathbf{x}, L^{\pi'_s, P_s} \rangle - r^{\pi'_s}, \\
\beta^{P_s} &\coloneqq r^{\pi''_s} - \langle \mathbf{x}, L^{\pi''_s, P_s} \rangle, \\
\mathscr{P}_s &\coloneqq \{P_s \mid \alpha^{P_s} \geq 0,\, \beta^{P_s} > 0,\, P_s \in \mathcal{P}_s\}, \\
\lambda &\coloneqq \min_{P_s \in \mathscr{P}_s} \frac{\alpha^{P_s}}{\alpha^{P_s} + \beta^{P_s}}.
\end{aligned} \tag{87}$$

and accordingly

$$1 - \lambda = \max_{P_s \in \mathscr{P}_s} \frac{\beta^{P_s}}{\alpha^{P_s} + \beta^{P_s}}. \tag{88}$$

We construct

$$\pi^\dagger_s \coloneqq (1-\lambda)\pi'_s + \lambda \pi''_s. \tag{89}$$

Note that $0 \leq \lambda \leq 1$. We have $\pi^\dagger_s \in \Delta_{\mathcal{A}}$ since $\pi^\dagger_s$ is a convex combination of π'_s and π''_s. Then we are going to show that $\mathbf{x} \in \mathcal{C}_+^{\pi^\dagger_s, \mathcal{P}_s} \cap \mathcal{C}_-^{\pi^\dagger_s, \mathcal{P}_s}$, *i.e.*,

$$\begin{aligned}
&\exists\, P_s \in \mathcal{P}_s, \quad \langle \mathbf{x}, L^{\pi^\dagger_s, P_s} \rangle - r^{\pi^\dagger_s} \geq 0; \\
&\forall\, P_s \in \mathcal{P}_s, \quad \langle \mathbf{x}, L^{\pi^\dagger_s, P_s} \rangle - r^{\pi^\dagger_s} \leq 0.
\end{aligned} \tag{90}$$

On the one hand, denoting

$$P^\dagger_s \coloneqq \operatorname*{arg\,min}_{P_s \in \mathscr{P}_s} \frac{\alpha^{P_s}}{\alpha^{P_s} + \beta^{P_s}}, \tag{91}$$

we have

$$\begin{aligned}
\langle \mathbf{x}, L^{\pi^\dagger_s, P^\dagger_s} \rangle - r^{\pi^\dagger_s} &= \langle \mathbf{x}, (1-\lambda) L^{\pi'_s, P^\dagger_s} + \lambda L^{\pi''_s, P^\dagger_s} \rangle - (1-\lambda) r^{\pi'_s} - \lambda r^{\pi''_s} \\
&= (1-\lambda) \left(\langle \mathbf{x}, L^{\pi'_s, P^\dagger_s} \rangle - r^{\pi'_s} \right) - \lambda \left(r^{\pi''_s} - \langle \mathbf{x}, L^{\pi''_s, P^\dagger_s} \rangle \right) \\
&= (1-\lambda) \alpha^{P^\dagger_s} - \lambda \beta^{P^\dagger_s} \\
&= \frac{\beta^{P^\dagger_s} \alpha^{P^\dagger_s}}{\alpha^{P^\dagger_s} + \beta^{P^\dagger_s}} - \frac{\alpha^{P\dagger_s} \beta^{P\dagger_s}}{\alpha^{P\dagger_s} + \beta^{P\dagger_s}} \\
&= 0.
\end{aligned} \tag{92}$$

On the other hand, for all $P_s \in \mathcal{P}_s$ we have

$$\begin{aligned}
\langle \mathbf{x}, L^{\pi^\dagger_s, P_s} \rangle - r^{\pi^\dagger_s} &= \langle \mathbf{x}, (1-\lambda) L^{\pi'_s, P_s} + \lambda L^{\pi''_s, P_s} \rangle - (1-\lambda) r^{\pi'_s} - \lambda r^{\pi''_s} \\
&= (1-\lambda) \left(\langle \mathbf{x}, L^{\pi'_s, P_s} \rangle - r^{\pi'_s} \right) - \lambda \left(r^{\pi''_s} - \langle \mathbf{x}, L^{\pi''_s, P_s} \rangle \right) \\
&= (1-\lambda) \alpha^{P_s} - \lambda \beta^{P_s} \\
&\leq \frac{\beta^{P_s} \alpha^{P_s}}{\alpha^{P_s} + \beta^{P_s}} - \lambda \beta^{P_s} \\
&\leq \frac{\beta^{P_s} \alpha^{P_s}}{\alpha^{P_s} + \beta^{P_s}} - \frac{\alpha^{P_s} \beta^{P_s}}{\alpha^{P_s} + \beta^{P_s}} \\
&\leq 0.
\end{aligned} \tag{93}$$

$$\begin{aligned}
\langle \mathbf{x}, L^{\pi_s^\dagger, P_s} \rangle - r^{\pi_s^\dagger} &= \langle \mathbf{x}, (1-\lambda) L^{\pi_s', P_s} + \lambda L^{\pi_s'', P_s} \rangle - (1-\lambda) r^{\pi_s'} - \lambda r^{\pi_s''} \\
&= (1-\lambda) \left(\langle \mathbf{x}, L^{\pi_s', P_s} \rangle - r^{\pi_s'} \right) - \lambda \left(r^{\pi_s''} - \langle \mathbf{x}, L^{\pi_s'', P_s} \rangle \right) \\
&= (1-\lambda)\alpha^{P_s} - \lambda \beta^{P_s} \\
&\leq \frac{\beta^{P_s} \alpha^{P_s}}{\alpha^{P_s} + \beta^{P_s}} - \lambda \beta^{P_s} \\
&\leq \frac{\beta^{P_s} \alpha^{P_s}}{\alpha^{P_s} + \beta^{P_s}} - \frac{\alpha^{P_s} \beta^{P_s}}{\alpha^{P_s} + \beta^{P_s}} \\
&\leq 0.
\end{aligned} \tag{94}$$

Putting it together, we obtain $\mathbf{x} \in \mathfrak{C}_+^{\pi_s^\dagger, \mathcal{P}_s} \cap \mathfrak{C}_-^{\pi_s^\dagger, \mathcal{P}_s}$ and thus $\mathbf{x} \in$ LHS. □

Lemma 4.7. *Consider an s-rectangular uncertainty set $\mathcal{P}$, we have for all states $s \in \mathcal{S}$,*

$$\bigcup_{\pi_s \in \Delta_\mathcal{A}} \mathfrak{C}_+^{\pi_s, \mathcal{P}_s} = \bigcup_{a \in \mathcal{A}} \mathfrak{C}_+^{d_{s,a}, \mathcal{P}_s}, \tag{32}$$

$$\bigcup_{\pi_s \in \Delta_\mathcal{A}} \mathfrak{C}_-^{\pi_s, \mathcal{P}_s} \supseteq \bigcup_{a \in \mathcal{A}} \mathfrak{C}_-^{d_{s,a}, \mathcal{P}_s}, \tag{33}$$

where the equality in the second line holds when $\mathcal{P}$ is (s,a)-rectangular.

Proof. **First**, we are going to prove

$$\bigcup_{\pi_s \in \Delta_\mathcal{A}} \mathfrak{C}_+^{\pi_s, \mathcal{P}_s} = \bigcup_{a \in \mathcal{A}} \mathfrak{C}_+^{d_{s,a}, \mathcal{P}_s}. \tag{95}$$

It is trivial that RHS $\subseteq$ LHS. We then focus on proving LHS $\subseteq$ RHS. For any $\mathbf{x} \in$ LHS, we have

$$\exists\, \pi_s' \in \Delta_\mathcal{A} \quad \text{s.t.} \quad \mathbf{x} \in \mathfrak{C}_+^{\pi_s', \mathcal{P}_s}. \tag{96}$$

Note that $\pi_s \in \Delta_\mathcal{A}$ can be written as a convex combination of $d_{s,a}, a \in \mathcal{A}$. In our case, we write

$$\pi_s' = \sum_{a \in \mathcal{A}} \pi_{s,a}' d_{s,a}. \tag{97}$$

Also note that for any $P_s \in \mathcal{P}_s$,

$$\langle \mathbf{x}, L^{\pi_s', P_s} \rangle - r^{\pi_s'} = \langle \mathbf{x}, \sum_{a \in \mathcal{A}} \pi_{s,a}' L^{d_{s,a}, P_s} \rangle - \sum_{a \in \mathcal{A}} \pi_{s,a}' r^{d_{s,a}} = \sum_{a \in \mathcal{A}} \pi_{s,a}' \left(\langle \mathbf{x}, L^{d_{s,a}, P_s} \rangle - r^{d_{s,a}} \right). \tag{98}$$

Therefore, we can write $\mathbf{x} \in \mathfrak{C}_+^{\pi_s', \mathcal{P}_s}$ as

$$\exists\, P_s \in \mathcal{P}_s, \quad \sum_{a \in \mathcal{A}} \pi_{s,a}' \left(\langle \mathbf{x}, L^{d_{s,a}, P_s} \rangle - r^{d_{s,a}} \right) \geq 0. \tag{99}$$

Since $\pi_{s,a}' \geq 0$ for all $a \in \mathcal{A}$, the above statement implies

$$\exists\, P_s \in \mathcal{P}_s, \exists\, a' \in \mathcal{A} \quad \text{s.t.} \quad \langle \mathbf{x}, L^{d_{s,a'}, P_s} \rangle - r^{d_{s,a'}} \geq 0. \tag{100}$$

This is equivalent to $\mathbf{x} \in$ RHS. Putting it together, we obtain LHS $=$ RHS.

Second, we are going to prove

$$\bigcup_{\pi_s \in \Delta_\mathcal{A}} \mathfrak{C}_-^{\pi_s, \mathcal{P}_s} \supseteq \bigcup_{a \in \mathcal{A}} \mathfrak{C}_-^{d_{s,a}, \mathcal{P}_s}, \tag{101}$$

where the equality holds when $\mathcal{P}$ is (s,a)-rectangular. Again, it is trivial that RHS $\subseteq$ LHS. We then focus on proving LHS $\subseteq$ RHS when $\mathcal{P}$ is (s,a)-rectangular. For any $\mathbf{x} \in$ LHS, we have

$$\exists\, \pi'_s \in \Delta_{\mathcal{A}} \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{C}_{-}^{\pi'_s, \mathcal{P}_s}. \tag{102}$$

Similarly, we can obtain

$$\forall\, P_s \in \mathcal{P}_s, \quad \sum_{a\in\mathcal{A}} \pi'_{s,a} \left(\langle \mathbf{x}, L^{d_{s,a}, P_s} \rangle - r^{d_{s,a}} \right) \leq 0. \tag{103}$$

This is equivalent to

$$\max_{P_s \in \mathcal{P}_s} \sum_{a\in\mathcal{A}} \pi'_{s,a} \left(\langle \mathbf{x}, L^{d_{s,a}, P_s} \rangle - r^{d_{s,a}} \right) \leq 0. \tag{104}$$

Due to (s,a)-rectangularity of $\mathcal{P}$, we have

$$\sum_{a\in\mathcal{A}} \pi'_{s,a} \max_{P_{s,a} \in \mathcal{P}_{s,a}} \left(\langle \mathbf{x}, L^{d_{s,a}, P_s} \rangle - r^{d_{s,a}} \right) \leq 0. \tag{105}$$

Since $\pi'_{s,a} \geq 0$ for all $a \in \mathcal{A}$, the above statement implies

$$\exists\, a' \in \mathcal{A}, \quad \text{s.t.} \quad \max_{P_{s,a'} \in \mathcal{P}_{s,a'}} \langle \mathbf{x}, L^{d_{s,a'}, P_s} \rangle - r^{d_{s,a'}} \leq 0, \tag{106}$$

which is equivalent to

$$\exists\, a' \in \mathcal{A},\, \forall\, P_{s,a'} \in \mathcal{P}_{s,a'} \quad \text{s.t.} \quad \langle \mathbf{x}, L^{d_{s,a'}, P_s} \rangle - r^{d_{s,a'}} \leq 0. \tag{107}$$

This is essentially saying $\mathbf{x} \in$ RHS. Putting it together, we obtain LHS $=$ RHS when $\mathcal{P}$ is (s,a)-rectangular. □

Theorem 4.8. *Consider an s-rectangular uncertainty set $\mathcal{P}$, the robust value function space $\mathcal{V}^{\mathcal{P}}$ satisfies*

$$\begin{aligned} \mathcal{V}^{\mathcal{P}} &= \bigcap_{s\in\mathcal{S}} \left[\left[\bigcup_{a\in\mathcal{A}} \mathcal{C}_{+}^{d_{s,a}, \mathcal{P}_s} \right] \cap \left[\bigcup_{\pi_s \in \Delta_{\mathcal{A}}} \mathcal{C}_{-}^{\pi_s, \mathcal{P}_s} \right] \right] \\ &\supseteq \bigcap_{s\in\mathcal{S}} \left[\left[\bigcup_{a\in\mathcal{A}} \mathcal{C}_{+}^{d_{s,a}, \mathcal{P}_s} \right] \cap \left[\bigcup_{a\in\mathcal{A}} \mathcal{C}_{-}^{d_{s,a}, \mathcal{P}_s} \right] \right] \end{aligned} \tag{34}$$

where the equality in the second line holds when $\mathcal{P}$ is (s,a)-rectangular.

Proof. The proof follows immediately from Lemma 4.5, Lemma 4.6 and Lemma 4.7. □

Lemma 4.9. *Consider a s-rectangular uncertainty set $\mathcal{P}$ and a policy π, we have*

$$(\mathcal{L}^{\pi_s, \mathcal{P}_s})^* = (g(\mathcal{P}_s))^* = (g(\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s))))^*. \tag{42}$$

Proof. Since affine transformations preserve affine hulls (Dattorro, 2005), we have

$$\begin{aligned} \mathbf{conv}(g(\mathcal{P}_s)) &= g(\mathbf{conv}(\mathcal{P}_s)), \\ \mathbf{conv}(g(\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s)))) &= g(\mathbf{conv}(\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s)))). \end{aligned} \tag{108}$$

Using Krein-Milman Theorem (Krein & Milman, 1940), we can obtain

$$g(\mathbf{conv}(\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s)))) = g(\mathbf{conv}(\mathcal{P}_s)). \tag{109}$$

Putting it together, we have

$$\mathbf{conv}(g(\mathcal{P}_s)) = \mathbf{conv}(g(\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s)))). \tag{110}$$

Then by the properties of polar cones (Proposition 2.2.1 in (Bertsekas, 2009)), we can get

$$(g(\mathcal{P}_s))^* = (g(\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s))))^*, \tag{111}$$

which completes the proof. □

Theorem 4.10. *Consider a s-rectangular uncertainty set $\mathcal{P}$, we have*

$$\mathcal{V}^{\mathcal{P}} = \mathcal{V}^{\mathcal{P}^\dagger} \tag{43}$$

where $\mathcal{P}^\dagger = \mathbf{ext}(\mathbf{conv}(\mathcal{P})) \subseteq \mathcal{P}$.

Proof. From Eqn. (41) and Lemma 4.9, we know that each conic hypersurface $C^{\pi_s,\mathcal{P}_s}$ only depends on $\mathbf{ext}(\mathbf{conv}(\mathcal{P}_s))$. Then we have

$$\mathcal{V}^{\mathcal{P}} = \mathcal{V}^{\mathcal{P}^\dagger}, \quad \text{where} \quad \mathcal{P}^\dagger = \bigtimes_{s\in\mathcal{S}} \mathbf{ext}(\mathbf{conv}(\mathcal{P}_s)). \tag{112}$$

By the definition of extreme points, it is straightforward to show that

$$\bigtimes_{s\in\mathcal{S}} \mathbf{ext}(\mathbf{conv}(\mathcal{P}_s)) = \mathbf{ext}\left(\bigtimes_{s\in\mathcal{S}} \mathbf{conv}(\mathcal{P}_s)\right). \tag{113}$$

Using the properties of Cartesian products (Bertsekas et al., 2003), we can get

$$\bigtimes_{s\in\mathcal{S}} \mathbf{conv}(\mathcal{P}_s) = \mathbf{conv}\left(\bigtimes_{s\in\mathcal{S}} \mathcal{P}_s\right) = \mathbf{conv}(\mathcal{P}). \tag{114}$$

Putting it together, we have $\mathcal{P}^\dagger = \mathbf{ext}(\mathbf{conv}(\mathcal{P}))$. Since $\mathcal{P}$ is assumed to be compact, then $\mathcal{P}^\dagger \subseteq \mathcal{P}$. □

Corollary 5.1. *Consider an s-rectangular uncertainty set $\mathcal{P}$, if an axis-parallel line intersects with the robust value space $\mathcal{V}^{\mathcal{P}}$, then the intersection will be a line segment.*

Proof. Without loss of generality, consider a line parallel to the axis corresponding to state s_1, and denote it as

$$K = \{\mathbf{u} + t\mathbf{e}_{s_1} \mid t \in \mathbb{R}\} \tag{115}$$

where $\mathbf{u} \in \mathbb{R}^{\mathcal{S}}$ is fixed. Then the intersection between this line and the robust value space is

$$K \cap \left[\bigcap_{s\in\mathcal{S}} \left[\left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_+^{\pi_s,\mathcal{P}_s}\right] \cap \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_-^{\pi_s,\mathcal{P}_s}\right]\right]\right]. \tag{116}$$

On the line K, denote the direction of the ray $\{\mathbf{u} + t\mathbf{e}_{s_1} \mid t \le 0\}$ as negative and the opposite direction as positive.

First, we have

$$K \cap \mathcal{H}_+^{\pi_s,P_s} = \left\{\mathbf{u} + t\mathbf{e}_{s_1} \mid t\langle \mathbf{e}_{s_1}, L^{\pi_s,P_s}\rangle \le r^{\pi_s} - \langle \mathbf{u}, L^{\pi_s,P_s}\rangle\right\} \tag{117}$$

For $s \neq s_1$, since $\langle \mathbf{e}_{s_1}, L^{\pi_s,P_s}\rangle \le 0$, the intersection $K \cap \mathcal{H}_+^{\pi_s,P_s}$ is either the line K or a negative ray. Thus, the intersection

$$K \cap \left[\bigcap_{s\in\mathcal{S}, s\neq s_1} \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_+^{\pi_s,\mathcal{P}_s}\right]\right] \tag{118}$$

is either the line K or a negative ray.

For $s = s_1$, since $\langle \mathbf{e}_{s_1}, L^{\pi_s,P_s}\rangle > 0$, then the intersection $K \cap \mathcal{H}_+^{\pi_s,P_s}$ is a positive ray. Thus, the intersection

$$K \cap \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_+^{\pi_s,\mathcal{P}_s}\right] \tag{119}$$

is also a positive ray.

Putting it together, we can obtain that the intersection

$$K \cap \left[\bigcap_{s\in\mathcal{S}} \left[\bigcup_{\pi_s\in\Delta_{\mathcal{A}}} \mathcal{C}_+^{\pi_s,\mathcal{P}_s}\right]\right] \tag{120}$$

is either empty or a line segment (or a point in degenerated case).

Similarly, we can show that the intersection

$$K \cap \left[\bigcap_{s \in \mathcal{S}} \left[\bigcup_{\pi_s \in \Delta_{\mathcal{A}}} \mathcal{C}_{-}^{\pi_s, \mathcal{P}_s} \right] \right] \tag{121}$$

is either empty or a line segment (or a point in degenerated case).

Finally, taking the intersection, we have that the intersection between K and the robust value space is either empty or a line segment (or a point in degenerated case). □

C. Additional Figures

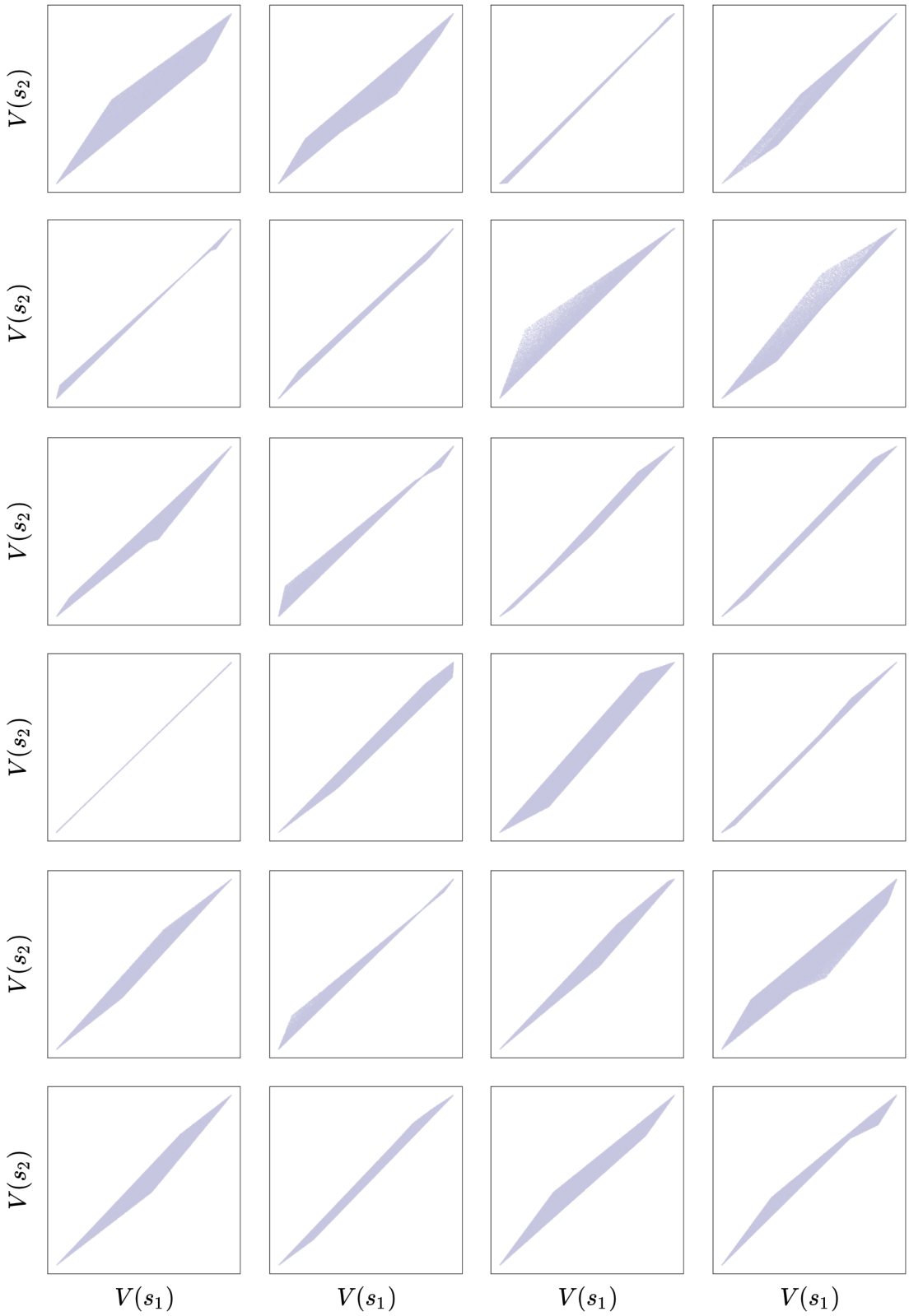

Figure 12. Visualization of the robust value space for several randomly generated s-rectangular RMDPs.