Appendix

A. Details of Markov Decision Processes

In this section we give the specifics of the Markov Decision Processes presented in this work. We will use the following convention:

$$r(s_i, a_j) = \hat{r}[i \times |\mathcal{A}| + j]$$
$$P(s_k|s_i, a_j) = \hat{P}[i \times |\mathcal{A}| + j][k]$$

where $\hat{P}, \hat{r}$ are the vectors given below.

In Section 3, Figure 2: (a) $|\mathcal{A}| = 2,\ \gamma = 0.9$
$\hat{r} = [0.06, 0.38, -0.13, 0.64]$
$\hat{P} = [[0.01, 0.99], [0.92, 0.08], [0.08, 0.92], [0.70, 0.30]]$

(b) $|\mathcal{A}| = 2,\ \gamma = 0.9$
$\hat{r} = [0.88, -0.02, -0.98, 0.42]$
$\hat{P} = [[0.96, 0.04], [0.19, 0.81], [0.43, 0.57], [0.72, 0.28]])$

(c) $|\mathcal{A}| = 3,\ \gamma = 0.9$
$\hat{r} = [-0.93, -0.49, 0.63, 0.78, 0.14, 0.41]$
$\hat{P} = [[0.52, 0.48], [0.5, 0.5], [0.99, 0.01], [0.85, 0.15], [0.11, 0.89], [0.1, 0.9]]$

(d) $|\mathcal{A}| = 2,\ \gamma = 0.9$
$\hat{r} = [-0.45, -0.1, 0.5, 0.5]$
$\hat{P} = [[0.7, 0.3], [0.99, 0.01], [0.2, 0.8], [0.99, 0.01]]$

In Section 3, Figure 3, 4, 5, 6: (left) $|\mathcal{A}| = 3,\ \gamma = 0.8$
$\hat{r} = [-0.1, -1., 0.1, 0.4, 1.5, 0.1]$
$\hat{P} = [[0.9, 0.1], [0.2, 0.8], [0.7, 0.3], [0.05, 0.95], [0.25, 0.75], [0.3, 0.7]]$

(right) $|\mathcal{A}| = 2,\ \gamma = 0.9$
$\hat{r} = [-0.45, -0.1, 0.5, 0.5]$
$\hat{P} = [[0.7, 0.3], [0.99, 0.01], [0.2, 0.8], [0.99, 0.01]]$

In Section 5: $|\mathcal{A}| = 2,\ \gamma = 0.9$
$\hat{r} = [-0.45, -0.1, 0.5, 0.5]$
$\hat{P} = [[0.7, 0.3], [0.99, 0.01], [0.2, 0.8], [0.99, 0.01]]$

B. Notation for the proofs

In the section we present the notation that we use to establish the results in the main text. The space of policies $\mathcal{P}(\mathcal{A})^{\mathcal{S}}$ describes a Cartesian product of simplices that we can express as a space of $|\mathcal{S}| \times |\mathcal{A}|$ matrices. However, we will adopt for policies, as well as the other components of $\mathcal{M}$, a convenient matrix form similar to (Wang et al., 2007).

- The transition matrix P is a $|\mathcal{S}||\mathcal{A}| \times |\mathcal{S}|$ matrix denoting the probability of going to state s' when taking action a in state s .
- A policy π is represented by a block diagonal $|\mathcal{S}| \times |\mathcal{S}||\mathcal{A}|$ matrix M_π. Suppose the state s is indexed by i and the action a is indexed by j in the matrix form, then we have that $M_\pi(i, i \times |\mathcal{A}| + j) = \pi(a|s)$. The rest of the entries of M_π are 0. From now on, we will confound π and M_π to enhance readability.
- The transition matrix $P^\pi = \pi P$ induced by a policy π is a $|\mathcal{S}| \times |\mathcal{S}|$ matrix denoting the probability of going from state s to state s' when following the policy π.
- The reward vector r is a $|\mathcal{S}||\mathcal{A}| \times 1$ matrix denoting the expected reward when taking action a in state s. The reward vector of a policy $r_\pi = \pi r$ is a $|\mathcal{S}| \times 1$ vector.
- The value function V^π of a policy π is a $|\mathcal{S}| \times 1$ matrix.
- We note C_i^π the i-th column of $(I - \gamma P^\pi)^{-1}$.

Under these notations, we can define the Bellman operator $\mathcal{T}^\pi$ and the optimality Bellman operator $\mathcal{T}^*$ as follows:

$$\mathcal{T}^\pi V^\pi = r_\pi + \gamma P^\pi V^\pi = \pi(r + \gamma P V^\pi)$$
$$\forall s \in \mathcal{S}, \mathcal{T}^* V^\pi(s) = \max_{\pi' \in \mathcal{P}(\mathcal{A})^{\mathcal{S}}} r_{\pi'}(s) + \gamma P^{\pi'} V^\pi(s).$$

C. Supplementary Results

Lemma 5. *f_v is infinitely differentiable on $\mathcal{P}(\mathcal{A})^{\mathcal{S}}$.*

Proof. We have that:

$$\begin{aligned} f_v(\pi) &= (I - \gamma\pi P)^{-1}\pi r \\ &= \frac{1}{\det(I - \gamma\pi P)} \text{adj}(I - \gamma\pi P)\pi r. \end{aligned}$$

Where det is the determinant and where adj is the adjunct. $\forall \pi \in \mathcal{P}(\mathcal{A})^{\mathcal{S}}, \det(I - \gamma\pi P) \neq 0$, therefore f_v is infinitely differentiable. □

Lemma 6. *Let $\pi \in \mathcal{P}(\mathcal{A})^{\mathcal{S}}$, $s_1, .., s_k \in \mathcal{S}$, and $\pi' \in Y^\pi_{s_1,..,s_k}$. We have*

$$Span(C^\pi_{k+1}, .., C^\pi_{|\mathcal{S}|}) = Span(C^{\pi'}_{k+1}, .., C^{\pi'}_{|\mathcal{S}|}).$$

Proof. As P^π and $P^{\pi'}$ are equal on their first k rows, we also have that $(I - \gamma P^\pi)$ and $(I - \gamma P^{\pi'})$ are equal on their first k rows. We note these k rows $L_1, ..., L_k$.

By assumption, we have that:

$$\forall i \in \{1, \ldots, k\}, \forall j \in \{k+1, \ldots, |\mathcal{S}|\}, L_i C^\pi_j = 0, L_i C^{\pi'}_j = 0.$$

Which we can rewrite,

$$Span(C^{\pi}_{k+1}, \dots, C^{\pi}_{|\mathcal{S}|}) \subset Span(L_1, \dots, L_k)^{\perp}$$
$$Span(C^{\pi'}_{k+1}, .., C^{\pi'}_{|\mathcal{S}|}) \subset Span(L_1, \dots, L_k)^{\perp}$$

Now using, $\dim Span(C^{\pi}_{k+1}, \dots, C^{\pi}_{|\mathcal{S}|}) = \dim Span(C^{\pi'}_{k+1}, .., C^{\pi'}_{|\mathcal{S}|}) = \dim Span(L_1, \dots, L_k)^{\perp} = |\mathcal{S}| - k$, we have:

$$Span(C^{\pi}_{k+1}, \dots, C^{\pi}_{|\mathcal{S}|}) = Span(L_1, \dots, L_k)^{\perp}$$
$$Span(C^{\pi'}_{k+1}, .., C^{\pi'}_{|\mathcal{S}|}) = Span(L_1, \dots, L_k)^{\perp}.$$

□

D. Proofs

Lemma 1. *The space of value functions $\mathcal{V}$ is compact and connected.*

Proof. $\mathcal{P}(\mathcal{A})^{\mathcal{S}}$ is connected since it is a convex space, and it is compact because it is closed and bounded in a finite dimensional real vector space. Since f_v is continuous (Lemma 5), we have $f_v(\mathcal{P}(\mathcal{A})^{\mathcal{S}}) = \mathcal{V}$ is compact and connected. □

Lemma 2. *Consider two policies π_1, π_2 that agree on $s_1, \dots, s_k \in \mathcal{S}$. Then the vector $r_{\pi_1} - r_{\pi_2}$ has zeros in the components corresponding to $s_1, \dots, s_k$ and the matrix $P^{\pi_1} - P^{\pi_2}$ has zeros in the corresponding rows.*

Proof. Suppose without loss of generality that $\{s_1, .., s_k\}$ are the first k states in the matrix form notation. We have,

$$r_{\pi_1} = \pi_1 r$$
$$r_{\pi_2} = \pi_2 r$$
$$P^{\pi_1} = \pi_1 P$$
$$P^{\pi_2} = \pi_2 P.$$

Since $\pi_1(\cdot \,|\, s) = \pi_2(\cdot \,|\, s)$ for all $s \in \{s_1, .., s_k\}$, the first k rows of π_1, π_2 are identical in the matrix form notation. Therefore, the first k elements of r_{π_1} and r_{π_2} are identical, and the first k rows of P^{π_1} and P^{π_2} are identical, hence the result. □

Lemma 3. *Consider a policy π and k states $s_1, \dots, s_k$. Then the value functions generated by $Y^{\pi}_{s_1,..,s_k}$ are contained in the affine vector space $H^{\pi}_{s_1,..,s_k}$:*

$$f_v(Y^{\pi}_{s_1,..,s_k}) = \mathcal{V} \cap H^{\pi}_{s_1,..,s_k}.$$

Proof. Let us first show that $f_v(Y^{\pi}_{s_1,..,s_k}) \subset \mathcal{V} \cap H^{\pi}_{s_1,..,s_k}$.
Let $\pi' \in Y^{\pi}_{s_1,..,s_k}$, i.e. π' agrees with π on $s_1, .., s_k$. Using Bellman's equation, we have:

$$\begin{aligned} V^{\pi'} - V^{\pi} &= r_{\pi'} - r_{\pi} + \gamma P^{\pi'} V^{\pi'} - \gamma P^{\pi} V^{\pi} \\ &= r_{\pi'} - r_{\pi} + \gamma (P^{\pi'} - P^{\pi}) V^{\pi'} + \gamma P^{\pi} (V^{\pi'} - V^{\pi}) \\ &= (I - \gamma P^{\pi})^{-1} \big(r_{\pi'} - r_{\pi} + \gamma (P^{\pi'} - P^{\pi}) V^{\pi'} \big). \end{aligned} \tag{3}$$

Since the policies π' and π agree on the states $s_1, \dots, s_k$, we have, using Lemma 2:

$$\begin{cases} r_{\pi'} - r_{\pi} & \text{is zero on its first k elements} \\ P^{\pi'} - P^{\pi} & \text{is zero on its first k rows.} \end{cases}$$

Hence, the right-hand side of Eq. 3 is the product of a matrix with a vector whose first k elements are 0. Therefore

$$V^{\pi'} \in V^{\pi} + Span(C^{\pi}_{k+1}, .., C^{\pi}_{|\mathcal{S}|}) \,.$$

We shall now show that $\mathcal{V} \cap H^{\pi}_{s_1,..,s_k} \subset f_v(Y^{\pi}_{s_1,..,s_k})$.

Suppose $V^{\hat{\pi}} \in H^{\pi}_{s_1,..,s_k}$. We want to show that there is a policy $\pi' \in Y^{\pi}_{s_1,..,s_k}$ such that $V^{\pi'} = V^{\hat{\pi}}$. We construct π' the following way:

$$\pi' = \begin{cases} \pi(\cdot \,|\, s) & \text{if } s \in \{s_1, .., s_k\} \\ \hat{\pi}(\cdot \,|\, s) & \text{otherwise.} \end{cases}$$

Therefore, using the result of the first implication of this proof:

$$V^{\hat{\pi}} - V^{\pi'} \in Span(C^{\pi}_{k+1}, .., C^{\pi}_{|\mathcal{S}|}) \text{ by assumption}$$
$$V^{\hat{\pi}} - V^{\pi'} \in Span(C^{\pi'}_{1}, .., C^{\pi'}_{k}) \text{ since } \hat{\pi} \text{ and } \pi' \text{ agree on } s_{k+1}, \ldots, s_{|\mathcal{S}|}.$$

However, as $\pi, \pi' \in Y^{\pi}_{s_1,..,s_k}$, we have using Lemma 6:

$$Span(C^{\pi}_{k+1}, .., C^{\pi}_{|\mathcal{S}|}) = Span(C^{\pi'}_{k+1}, .., C^{\pi'}_{|\mathcal{S}|}).$$

Therefore, $V^{\hat{\pi}} - V^{\pi'} \in Span(C^{\pi'}_{1}, .., C^{\pi'}_{k}) \cap Span(C^{\pi'}_{k+1}, .., C^{\pi'}_{|\mathcal{S}|}) = \{0\}$, meaning that $V^{\hat{\pi}} = V^{\pi'} \in f_v(Y^{\pi}_{s_1,..,s_k})$. □

Lemma 4. *Consider the ensemble $Y^{\pi}_{\mathcal{S}\setminus\{s\}}$ of policies that agree with a policy π everywhere but on $s \in \mathcal{S}$. For $\pi_0, \pi_1 \in Y^{\pi}_{\mathcal{S}\setminus\{s\}}$ define the function $g : [0,1] \to \mathcal{V}$*

$$g(\mu) = f_v(\mu\pi_1 + (1-\mu)\pi_0).$$

Then the following hold regarding g:

(i) g is continuously differentiable;

(ii) (Total order) $g(0) \preccurlyeq g(1)$ or $g(0) \succcurlyeq g(1)$;

(iii) If $g(0) = g(1)$ then $g(\mu) = g(0)$, $\mu \in [0,1]$;

(iv) (Monotone interpolation) If $g(0) \neq g(1)$ there is a $\rho : [0,1] \to \mathbb{R}$ such that $g(\mu) = \rho(\mu)g(1) + (1-\rho(\mu))g(0)$, and ρ is a strictly monotonic rational function of μ.

Proof. (i) g is continuously differentiable as a composition of two continuously differentiable functions.

(ii) We want to show that we have either $V^{\pi_1} \preccurlyeq V^{\pi_0}$ or $V^{\pi_1} \succcurlyeq V^{\pi_0}$.

Suppose, without loss of generality, that s is the first state in the matrix form. Using Lemma 3, we have:

$$V^{\pi_0} = V^{\pi_1} + \alpha C^{\pi_1}_1, \text{ with } \alpha \in \mathbb{R}.$$

As $(I - \gamma P^{\pi_1})^{-1} = \sum_{i=0}^{\infty} (\gamma \pi P)^i$, whose entries are all positive, $C^{\pi_1}_1$ is a vector with positive entries. Therefore we have $V^{\pi_1} \preccurlyeq V^{\pi_0}$ or $V^{\pi_1} \succcurlyeq V^{\pi_0}$, depending on the sign of α.

(iii) We have, using Equation (3)

$$V^{\pi_0} - V^{\pi_\mu} = (I - \gamma P^{\pi_\mu})^{-1}\big(r_{\pi_0} - r_{\pi_\mu} + \gamma(P^{\pi_0} - P^{\pi_\mu})V^{\pi_0}\big)$$
$$V^{\pi_0} - V^{\pi_1} = (I - \gamma P^{\pi_1})^{-1}\big(r_{\pi_0} - r_{\pi_1} + \gamma(P^{\pi_0} - P^{\pi_1})V^{\pi_0}\big).$$

Now, using

$$\begin{aligned} r_{\pi_0} &= \pi_0 r \\ r_{\pi_\mu} &= \pi_\mu r = \pi_0 r + \mu(\pi_1 - \pi_0) r \\ P^{\pi_0} &= \pi_0 P \\ P^{\pi_\mu} &= \pi_\mu P = \pi_0 P + \mu(\pi_1 - \pi_0) P \ , \end{aligned}$$

we have

$$\begin{aligned} V^{\pi_0} - V^{\pi_\mu} &= \mu(I - \gamma P^{\pi_\mu})^{-1}\big(r_{\pi_0} - r_{\pi_1} + \gamma(P^{\pi_0} - P^{\pi_1})V^{\pi_0}\big) \\ &= \mu(I - \gamma P^{\pi_\mu})^{-1}(I - \gamma P^{\pi_1})(V^{\pi_0} - V^{\pi_1}). \end{aligned} \tag{4}$$

Therefore, $g(0) = g(1) \Rightarrow V^{\pi_0} - V^{\pi_1} = 0 \Rightarrow V^{\pi_0} - V^{\pi_\mu} = 0 \Rightarrow g(\mu) = g(0)$.

(iv) If $g(0) = g(1)$, the result is true since we can take $\rho = 0$ using (iii).

Suppose $g(0) \neq g(1)$, let us prove the existence of ρ and that it is a rational function in μ. Reusing the Equation 4, we have

$$\begin{aligned} V^{\pi_0} - V^{\pi_\mu} &= \mu(I - \gamma P^{\pi_\mu})^{-1}(I - \gamma P^{\pi_1})(V^{\pi_0} - V^{\pi_1}) \\ &= \mu(I - \gamma(P^{\pi_0} + \mu(P^{\pi_1} - P^{\pi_0})))^{-1}(I - \gamma P^{\pi_1})(V^{\pi_0} - V^{\pi_1}) \\ &= \mu(I - \gamma P^{\pi_1} - \gamma(1-\mu)(P^{\pi_0} - P^{\pi_1})))^{-1}(I - \gamma P^{\pi_1})(V^{\pi_0} - V^{\pi_1}). \end{aligned}$$

As we have that $P^{\pi_0} - P^{\pi_1}$ is a rank one matrix (Lemma 2) that we can express as $P^{\pi_0} - P^{\pi_1} = uv^t$ with $u, v \in \mathbb{R}^{|\mathcal{S}|}$. From the Sherman-Morrison formula:

$$(I - \gamma P^{\pi_1} - \gamma(1-\mu)(P^{\pi_0} - P^{\pi_1})))^{-1} = (I - \gamma P^{\pi_1})^{-1} - \gamma(1-\mu)\frac{(I - \gamma P^{\pi_1})^{-1}(P^{\pi_0} - P^{\pi_1})(I - \gamma P^{\pi_1})^{-1}}{1 + \gamma(1-\mu)v^t(I - \gamma P^{\pi_1})^{-1}u}.$$

Define $\omega_{\pi_1,\pi_0} = v^t(I - \gamma P^{\pi_1})^{-1}u$, we have

$$V^{\pi_0} - V^{\pi_\mu} = \mu V^{\pi_0} - \mu V^{\pi_1} - \frac{\gamma\mu(1-\mu)}{1 + \omega_{\pi_1,\pi_0}\gamma(1-\mu)}(I - \gamma P^{\pi_1})^{-1}(P^{\pi_0} - P^{\pi_1})(V^{\pi_0} - V^{\pi_1}).$$

As in (i) we have that $(P^{\pi_0} - P^{\pi_1})(V^{\pi_0} - V^{\pi_1})$ is zeros on its last $\mathcal{S} - 1$ elements using an argument similar to Lemma 2, therefore

$$(I - \gamma P^{\pi_1})^{-1}(P^{\pi_0} - P^{\pi_1})(V^{\pi_0} - V^{\pi_1}) = \beta_{\pi_0,\pi_1} C_1^{\pi_1}, \text{ with } \beta_{\pi_0,\pi_1} \in \mathbb{R}.$$

Now recall from the proof of (i) that similarly, we have

$$V^{\pi_0} - V^{\pi_1} = \alpha_{\pi_0,\pi_1} C_1^{\pi_1}.$$

As by assumption $V^{\pi_0} - V^{\pi_1} \neq 0$, we have:

$$(I - \gamma P^{\pi_1})^{-1}(P^{\pi_0} - P^{\pi_1})(V^{\pi_0} - V^{\pi_1}) = \frac{\beta_{\pi_0,\pi_1}}{\alpha_{\pi_0,\pi_1}}(V^{\pi_0} - V^{\pi_1}).$$

Finally we have,

$$\begin{aligned} V^{\pi_0} - V^{\pi_\mu} &= \mu V^{\pi_0} - \mu V^{\pi_1} - \frac{\gamma\mu(1-\mu)}{1 + \omega_{\pi_1,\pi_0}\gamma(1-\mu)}\frac{\beta_{\pi_0,\pi_1}}{\alpha_{\pi_0,\pi_1}}(V^{\pi_0} - V^{\pi_1}) \\ &= \Big(\mu - \frac{\gamma\mu(1-\mu)}{1 + \omega_{\pi_1,\pi_0}\gamma(1-\mu)}\frac{\beta_{\pi_0,\pi_1}}{\alpha_{\pi_0,\pi_1}}\Big)(V^{\pi_0} - V^{\pi_1}). \end{aligned}$$

Therefore, ρ is a rational function in μ, hence continuous, that we can express as:

$$\rho(\mu) = \mu - \frac{\gamma\mu(1-\mu)}{1+\omega_{\pi_1,\pi_0}\gamma(1-\mu)}\frac{\beta_{\pi_0,\pi_1}}{\alpha_{\pi_0,\pi_1}}.$$

Now let us prove that ρ is strictly monotonic. Suppose that ρ is not strictly monotonic. As ρ is continuous, we have that ρ is not injective. Hence, $\exists \mu_0, \mu_1 \in [0,1]$ distinct and the associated mixture of policies π_{μ_0}, π_{μ_1} such that

$$\begin{aligned}
g(\mu_0) = g(\mu_1) &\Leftrightarrow V^{\pi_{\mu_0}} = V^{\pi_{\mu_1}} \\
&\Leftrightarrow \mathcal{T}^{\pi_{\mu_0}} V^{\pi_{\mu_0}} = \mathcal{T}^{\pi_{\mu_1}} V^{\pi_{\mu_1}} \\
&\Leftrightarrow \mathcal{T}^{\pi_{\mu_0}} V^{\pi_{\mu_0}} = \mathcal{T}^{\pi_{\mu_1}} V^{\pi_{\mu_0}} \\
&\Leftrightarrow (\mu_0\mathcal{T}^{\pi_0} + (1-\mu_0)\mathcal{T}^{\pi_1})V^{\pi_{\mu_0}} = (\mu_1\mathcal{T}^{\pi_0} + (1-\mu_1)\mathcal{T}^{\pi_1})V^{\pi_{\mu_0}} \\
&\Leftrightarrow \mu_0(\mathcal{T}^{\pi_0} - \mathcal{T}^{\pi_1})V^{\pi_{\mu_0}} = \mu_1(\mathcal{T}^{\pi_0} - \mathcal{T}^{\pi_1})V^{\pi_{\mu_0}} \\
&\Leftrightarrow \mathcal{T}^{\pi_0} V^{\pi_{\mu_0}} = \mathcal{T}^{\pi_1} V^{\pi_{\mu_0}}.
\end{aligned}$$

Therefore we have

$$V^{\pi_{\mu_0}} = \mathcal{T}^{\pi_{\mu_0}} V^{\pi_{\mu_0}} = (\mu_0\mathcal{T}^{\pi_0} + (1-\mu_0)\mathcal{T}^{\pi_1})V^{\pi_{\mu_0}} = \mathcal{T}^{\pi_1} V^{\pi_{\mu_0}}.$$

Therefore

$$\mathcal{T}^{\pi_0} V^{\pi_{\mu_0}} = \mathcal{T}^{\pi_1} V^{\pi_{\mu_0}} = V^{\pi_{\mu_0}}.$$

However, the Bellman operator has a unique fixed point, therefore

$$V^{\pi_0} = V^{\pi_1} = V^{\pi_{\mu_0}},$$

which contradicts our assumption. □

Theorem 1. *[Line Theorem] Let s be a state and π, a policy. Then there are two s-deterministic policies in $Y^{\pi}_{\mathcal{S}\setminus\{s\}}$, denoted π_l, π_u, which bracket the value of all other policies $\pi' \in Y^{\pi}_{\mathcal{S}\setminus\{s\}}$:*

$$f_v(\pi_l) \preccurlyeq f_v(\pi') \preccurlyeq f_v(\pi_u).$$

Furthermore, the image of f_v restricted to $Y^{\pi}_{\mathcal{S}\setminus\{s\}}$ is a line segment, and the following three sets are equivalent:

(i) $f_v\big(Y^{\pi}_{\mathcal{S}\setminus\{s\}}\big)$,

(ii) $\{f_v(\alpha\pi_l + (1-\alpha)\pi_u) \,|\, \alpha \in [0,1]\}$,

(iii) $\{\alpha f_v(\pi_l) + (1-\alpha) f_v(\pi_u) \,|\, \alpha \in [0,1]\}$.

Proof. Let us start by proving the first statement of the theorem which is the existence of two s-deterministic policies π_u, π_l in $Y^{\pi}_{\mathcal{S}\setminus\{s\}}$ that respectively dominates and is dominated by all other policies.

The existence of π_l and π_u (without enforcing their s-determinism), whose value functions are respectively dominated or dominate all other value functions of policies of $Y^{\pi}_{\mathcal{S}\setminus\{s\}}$, is given by:

- $f_v(Y^{\pi}_{\mathcal{S}\setminus\{s\}})$ is compact as an intersection of a compact and an affine plane (Lemma 3).
- There is a total order on this compact space ((ii) in Lemma 4).

Suppose π_l is not s-deterministic, then there is $a \in \mathcal{A}$ such that $\pi_l(a|s) = \mu^* \in (0,1)$. Hence we can write π_l as a mixture of π_1, π_2 defined as follows

$$\forall s' \in \mathcal{S}, a' \in \mathcal{A}, \pi_1(a'|s') = \begin{cases} 1 \text{ if } s' = s, a' = a \\ 0 \text{ if } s' = s, a' \neq a \\ \pi_l(a'|s') \text{ otherwise.} \end{cases} \tag{5}$$

$$\pi_2(a'|s') = \begin{cases} 0 \text{ if } s' = s, a' = a \\ \frac{1}{1-\mu^*}\pi_l(a'|s') \text{ if } s' = s, a' \neq a \\ \pi_l(a'|s') \text{ otherwise.} \end{cases} \tag{6}$$

Therefore $\pi_l = \mu^*\pi_1 + (1-\mu^*)\pi_2$. We can use (iv) in Lemma 4, that gives that $g : \mu \mapsto f_v(\mu\pi_1 + (1-\mu)\pi_2)$ is either strictly monotonic or constant. If g was strictly monotonic we would have a contradiction on $f_v(\pi_l)$ being minimum. Therefore g is constant, and in particular

$$f_v(\pi_l) = g(1) = f_v(\pi_1),$$

with π_1 s-deterministic.

Similarly we can show there is an s-deterministic policy that has the same value function as π_u, hence proving the result.

Now let us prove the equivalence between (i), (ii) and (iii).

- Let $\pi' \in Y^{\pi}_{\mathcal{S}\setminus\{s\}}$, we have: $f_v(\pi_l) \preccurlyeq f_v(\pi') \preccurlyeq f_v(\pi_u)$ and $f_v(\pi_l), f_v(\pi'), f_v(\pi_u)$ are on the same line (Lemma 3). Therefore, $f_v(\pi')$ is a convex combination of $f_v(\pi_l)$ and $f_v(\pi_u)$ hence (i) $\subset$ (iii).
- By definition, (ii) $\subset$ (i).
- Lemma 4 gives $\mu \mapsto f_v(\mu\pi_u + (1-\mu)\pi_l) = f_v(\pi_l) + \rho(\mu)(f_v(\pi_u) - f_v(\pi_l))$ with ρ continuous and $\rho(0) = 0, \rho(1) = 1$. Using the theorem of intermediary values on ρ we have that ρ takes all values between 0 and 1. Therefore (iii) $\subset$ (ii)

We have (iii) $\subset$ (ii) $\subset$ (i) $\subset$ (iii). Therefore (i) = (ii) = (iii).

□

Corollary 1. *For any set of states $s_1, .., s_k \in \mathcal{S}$ and a policy π, V^π can be expressed as a convex combination of value functions of $\{s_1, .., s_k\}$-deterministic policies. In particular, $\mathcal{V}$ is included in the convex hull of the value functions of deterministic policies.*

Proof. We prove the result by induction on the number of states k. If $k = 1$, the result is true by Theorem 1.

Suppose the result is true for k states. Let $s_1, .., s_{k+1} \in \mathcal{S}$, we have by assumption that

$$\exists n \in \mathbb{N}, \pi_1, .., \pi_n \{s_1, .., s_k\}\text{-deterministic} \in \mathcal{P}(\mathcal{A})^{\mathcal{S}}, \alpha_1, .., \alpha_n \in [0,1], \text{ s.t. } \begin{cases} V^\pi = \sum_{i=1}^n \alpha_i V^{\pi_i} \\ \sum_{i=1}^n \alpha_i = 1 \end{cases}$$

However, using Theorem 1, we have

$$\forall i \in [1,n], \exists \pi_{i,l}, \pi_{i,u} \in \mathcal{P}(\mathcal{A})^{\mathcal{S}}, \begin{cases} \pi_{i,l}, \pi_{i,u} \ s_{k+1}\text{-deterministic} \\ \pi_{i,l}, \pi_{i,u} \text{ agrees with } \pi_i \text{ on } s_1, .., s_k \\ \exists \beta_i \in [0,1], V^{\pi_i} = \beta_i V^{\pi_{i,l}} + (1-\beta_i) V^{\pi_{i,u}} \end{cases}.$$

Therefore

$$V^\pi = \sum_{i=1}^n \alpha_i (\beta_i V^{\pi_{i,l}} + (1-\beta_i) V^{\pi_{i,u}}),$$

thus concluding the proof. □

Proposition 1. *P is a polyhedron in an affine subspace $K \subseteq \mathbb{R}^n$ if*

(i) P *is closed;*

(ii) There are $k \in \mathbb{N}$ *hyperplanes* $H_1, .., H_k$ *in* K *whose union contains the boundary of* P *in* K*:* $\partial_K P \subset \cup_{i=1}^k H_i$*; and*

(iii) For each of these hyperplanes, $P \cap H_i$ *is a polyhedron in* H_i*.*

Proof. We will show the result by induction on the dimension of K.

For $\dim(K) = 1$, the proposition is true since P is a polyhedron iff its boundary is a finite number of points.

Suppose the proposition is true for $\dim(K) = n$, let us show that it is true for $\dim(K) = n + 1$.

We can verify that if P is a polyhedron, then:

- P is closed.
- There is a finite number of hyperplanes covering its boundaries (the boundaries of the half-spaces defining each convex polyhedron composing P).
- The intersection of P with these hyperplanes still are polyhedra.

Now let us consider the other direction of the implication, i.e. suppose that P is closed, $\partial_K P \subset \cap_{i=1}^k H_i$, and $\forall i, P \cap H_i$ is a polyhedron. We will show that we can express P as a finite union of polyhedra.

Suppose $x \in P$ and $x \notin \cup_{i=1}^k H_i$. By assumption, we have that $x \in \text{relint}_K(P)$. We will show that x is in a intersection of closed half-spaces defined by the hyperplanes $H_1, .., H_k$ and that any other vector in this intersection is also in P (otherwise we would have a contradiction on the boundary assumption).

A hyperplane H_i defines two closed half-spaces denoted by H_i^{+1} and H_i^{-1} (the signs being arbitrary). And the intersections of those half-spaces form a partition of K, therefore:

$$\exists \delta \in \{-1, 1\}^k, x \in \cap_{i=1}^k H_i^{\delta(i)} = P_\delta.$$

By assumption, $x \in \text{relint}_K(P_\delta)$, since we assumed that $x \notin \cup_{i=1}^k H_i$. Now suppose $\exists y \in \text{relint}_K(P_\delta)$ s.t. $y \notin P$, we have:

$$\exists \lambda \in [0, 1], \lambda x + (1 - \lambda)y = z \in \partial_K P.$$

However, $z \in \text{relint}_K(P_\delta)$ because $\text{relint}_K(P_\delta)$ is convex. Therefore $z \notin \bigcup_{i=1}^k H_i$ since $z \in \text{relint}_K(P_\delta)$ which gives a contradiction. We thus have either $\text{relint}_K(P_\delta) \subset P$ or $\text{relint}_K(P_\delta) \cap P = \emptyset$.

Now suppose $\text{relint}_K(P_\delta) \subset P$ and $\text{relint}_K(P_\delta)$ nonempty. We have that $\text{cl}_K(\text{relint}_K(P_\delta)) = P_\delta$ (Brondsted, 2012, Theorem 3.3) and P closed, meaning that $P_\delta \subset P$. Therefore, we have

$$\exists \delta_1, .., \delta_j \in \{-1, 1\}^k \text{ s.t. } P = (\cup_{i=1}^k P \cap H_i) \bigcup (\cup_{i=1}^j P_{\delta_i}).$$

P is thus a finite union polyhedra, as $\{P \cap H_i\}$ are polyhedra by assumption, and $\{P_{\delta_i}\}$ are convex polyhedra by definition. □

Corollary 2. *Let* V^π *and* $V^{\pi'}$ *be two value functions. Then there exists a sequence of* $k \leq |\mathcal{S}|$ *policies,* $\pi_1, \ldots, \pi_k$*, such that* $V^\pi = V^{\pi_1}$*,* $V^{\pi'} = V^{\pi_k}$*, and for every* $i \in 1, \ldots, k-1$*, the set*

$$\{f_v(\alpha\pi_i + (1 - \alpha)\pi_{i+1}) \,|\, \alpha \in [0, 1]\}$$

forms a line segment.

Proof. We can define the policies $\pi_2, .., \pi_{|\mathcal{S}|-1}$ the following way:

$$\forall i \in [2, |\mathcal{S}| - 1], \begin{cases} \pi_i(\cdot \mid s_j) = \pi'(\cdot \mid s_j) & \text{if } s_j \in \{s_1, .., s_{i-1}\} \\ \pi_i(\cdot \mid s_j) = \pi(\cdot \mid s_j) & \text{if } s_j \in \{s_i, .., s_{|\mathcal{S}|}\} \end{cases}$$

Therefore, two consecutive policies π_i, π_{i+1} only differ on one state. We can apply Theorem 1 and thus conclude the proof. □

Theorem 2. *Consider the ensemble of policies $Y^{\pi}_{s_1,..,s_k}$ that agree with π on states $\mathscr{S} = \{s_1, .., s_k\}$. Suppose $\forall s \notin \mathscr{S}$, $\forall a \in \mathcal{A}$, $\nexists \pi' \in Y^{\pi}_{s_1,..,s_k} \cap D_{a,s}$ s.t. $f_v(\pi') = f_v(\pi)$, then $f_v(\pi)$ has a relative neighborhood in $\mathcal{V} \cap H^{\pi}_{s_1,..,s_k}$.*

Proof. We will prove the result by showing that V^{π} is in $|\mathcal{S}| - k$ segments that are linearly independent by applying the line theorem on a policy $\hat{\pi}$ that has the same value function as π. We will then be able to conclude using the regularity of f_v.

We can find a policy $\hat{\pi} \in Y^{\pi}_{s_1,..,s_k}$, that has the same value function as π by applying recursively Theorem 1 on the states $\{s_{k+1}, .., s_{|\mathcal{S}|}\}$, such that:

$\exists a_{k+1,l}, a_{k+1,u}, .., a_{|\mathcal{S}|,l}, a_{|\mathcal{S}|,u} \in \mathcal{A}, \forall i \in \{k+1, .., |\mathcal{S}|\}$,

$$\hat{\pi}(a_{i,l}|s_i) = 1 - \hat{\pi}(a_{i,u}|s_i) = \hat{\mu}_i \in (0, 1).$$

Note that $\hat{\mu}_i \notin \{0, 1\}$ because we assumed that no s-deterministic policy has the same value function as π.
We define $\hat{\mu} = (\hat{\mu}_{k+1}, ..., \hat{\mu}_{|\mathcal{S}|}) \in (0, 1)^{|\mathcal{S}|-k}$ and the function $g : (0, 1)^{|\mathcal{S}|-k} \to H^{\pi}_{s_1,..,s_k}$ such that:

$$g(\mu) = f_v(\pi_\mu), \text{ with } \begin{cases} \pi_\mu(a_{i,l}|s_i) = 1 - \pi_\mu(a_{i,u}|s_i) = \mu_i & \text{if } i \in \{k+1, \ldots, |\mathcal{S}|\} \\ \pi_\mu(\cdot \mid s_i) = \hat{\pi}(\cdot \mid s_i) & \text{otherwise.} \end{cases}$$

We have that:

1. g is continuously differentiable
2. $g(\hat{\mu}) = f_v(\hat{\pi})$
3. $\frac{\partial g}{\partial \mu_i}$ is non-zero at $\hat{\mu}$ (Lemma 4.iv)
4. $\frac{\partial g}{\partial \mu_i}$ is along the i-th column of $(I - \gamma P^{\hat{\pi}})^{-1}$ (Lemma 3)

Therefore, the Jacobian of g is invertible at μ since the columns of $(I - \gamma P^{\hat{\pi}})^{-1}$ are independent, therefore by the inverse theorem function, there is a neighborhood of $g(\mu) = f_v(\hat{\pi})$ in the image space, which gives the result. □

Corollary 3. *Consider a policy $\pi \in \mathcal{P}(\mathcal{A})^{\mathcal{S}}$, the states $\mathscr{S} = \{s_1, .., s_k\}$, and the ensemble $Y^{\pi}_{s_1,..,s_k}$ of policies that agree with π on $s_1, .., s_k$. Define $\mathcal{V}^y = f_v(Y^{\pi}_{s_1,..,s_k})$, we have that the relative boundary of $\mathcal{V}^y$ in $H^{\pi}_{s_1,..,s_k}$ is included in the value functions spanned by policies in $Y^{\pi}_{s_1,..,s_k}$ that are s-deterministic for $s \notin \mathscr{S}$:*

$$\partial \mathcal{V}^y \subset \bigcup_{s \notin \mathscr{S}} \bigcup_{a \in \mathcal{A}} f_v(Y^{\pi}_{s_1,..,s_k} \cap D_{s,a}),$$

where ∂ refers to $\partial_{H^{\pi}_{s_1,..,s_k}}$.

Proof. Let $V^{\pi} \in \mathcal{V}^y$; from Theorem 2, we have that

$$V^{\pi} \notin \bigcup_{i=k+1}^{|\mathcal{S}|} \bigcup_{j=1}^{|\mathcal{A}|} f_v(Y^{\pi}_{s_1,..,s_k} \cap D_{s_i,a_j}) \Rightarrow V^{\pi} \in \text{relint}(V^y)$$

where relint refers to the relative interior in $H^{\pi}_{s_1,..,s_k}$.

Therefore,

$$\partial \mathcal{V}^y \subset \bigcup_{i=k+1}^{|\mathcal{S}|} \bigcup_{j=1}^{|\mathcal{A}|} f_v(Y^{\pi}_{s_1,..,s_k} \cap D_{s_i,a_j}).$$

□

Theorem 3. *Consider a policy $\pi \in \mathcal{P}(\mathcal{A})^{\mathcal{S}}$, the states $s_1, .., s_k \in \mathcal{S}$, and the ensemble $Y^{\pi}_{s_1,..,s_k}$ of policies that agree with π on $s_1, .., s_k$. Then $f_v(Y^{\pi}_{s_1,..,s_k})$ is a polytope and in particular, $\mathcal{V} = f_v(Y^{\pi}_{\emptyset})$ is a polytope.*

Proof. We prove the result by induction on the cardinality of the number of states k.
If $k = |\mathcal{S}|$, then $Y^{\pi}_{s_1,..,s_k} = \{f_v(\pi)\}$ which is a polytope.

Suppose that the result is true for $k+1$, let us show that it is still true for k.

Let $\pi \in \Pi, s_1, .., s_k \in \mathcal{S}$, define the ensemble $Y^{\pi}_{s_1,..,s_k}$ of policies that agree with π on $\{s_1, .., s_k\}$ and $\mathcal{V}^y = f_v(Y^{\pi}_{s_1,..,s_k})$. Using Lemma 3, we have that $\mathcal{V}^y = \mathcal{V} \cap H^{\pi}_{s_1,..,s_k}$.

Now, using Corollary 3, we have that:

$$\partial \mathcal{V}^y \subset \bigcup_{i=k+1}^{|\mathcal{S}|} \bigcup_{j=1}^{|\mathcal{A}|} f_v(Y^{\pi}_{s_1,..,s_k} \cap D_{s_i,a_j}) = \bigcup_{i=k+1}^{|\mathcal{S}|} \bigcup_{j=1}^{|\mathcal{A}|} \mathcal{V}^y \cap H_{i,j},$$

where ∂ refer to the relative boundary in $H^{\pi}_{s_1,..,s_k}$, and $H_{i,j}$ is an affine hyperplane of $H^{\pi}_{s_1,..,s_k}$ (Lemma 3).

Therefore we have that:

1. $\mathcal{V}^y = \mathcal{V} \cap H^{\pi}_{s_1,..,s_k}$ is closed since it is an intersection of two closed ensembles
2. $\partial \mathcal{V}^y \subset \bigcup_{i=k+1}^{|\mathcal{S}|} \bigcup_{j=1}^{|\mathcal{A}|} H_{i,j}$ affine hyperplanes in $H^{\pi}_{s_1,..,s_k}$
3. $\mathcal{V}^y \cap H_{i,j} = f_v(Y^{\pi}_{s_1,..,s_k} \cap D_{s_i,a_j})$ is a polyhedron (induction assumption).

$\mathcal{V}^y$ verifies (i), (ii), (iii) in Proposition 1, therefore it is a polyhedron. We have $\mathcal{V}^y$ bounded since $\mathcal{V}^y \subset \mathcal{V}$ bounded. Therefore, $\mathcal{V}^y$ is a polytope. □