Supplementary Materials
Hilbert Space Embeddings of POMDPs

1 Supplementary Materials

In kernel methods, Bellman operators are considered with the signed measures $\boldsymbol{\alpha}, \boldsymbol{\beta'}_{a;\boldsymbol{\alpha}}, \boldsymbol{\alpha}'_{a;\boldsymbol{\alpha}}$ on finite sample sets $\mathcal{S}_0$, $\mathcal{O}_0$ instead of sets $\mathcal{S}$, $\mathcal{O}$. Let $\hat{H}_n$ be the Bellman operator

$$\left(\hat{H}_n V\right)(\boldsymbol{\alpha}) = \max_{a\in\mathcal{A}} \left[\boldsymbol{\alpha}^\top \boldsymbol{R}_a + \gamma \boldsymbol{\beta'}^\top_{a;\boldsymbol{\alpha}} \mathbf{V}\left(\boldsymbol{\alpha}'_{a,\mathcal{O}_0}\right)\right]. \quad (1)$$

and $\hat{H}_n^+$ be the corrected Bellman operator using probability vectors $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta'}}_{a;\boldsymbol{\alpha}}, \hat{\boldsymbol{\alpha}'}_{a;\boldsymbol{\alpha}}$. Consider a set $\mathcal{I}(\subset \mathbb{R}^n)$ such that $\boldsymbol{\alpha}'_{a,\mathcal{O}_0} \in \mathcal{I}$ for all $\boldsymbol{\alpha} \in \mathcal{I}$.

Theorem 1. *If $\boldsymbol{\beta'}_{a;\boldsymbol{\alpha}} \geq \mathbf{0}$ for all $\boldsymbol{\alpha} \in \mathcal{I}$ and $a \in \mathcal{A}$, kernel Bellman operator $\hat{H}_n$ is isotonic in $\mathcal{I}$, i.e., for any two value functions V and W, if $V \leq W$ in $\mathcal{I}$, $\hat{H}_n V \leq \hat{H}_n W$ in $\mathcal{I}$.*

Proof. The proof is similar to [Porta. et al., 2006]. Let $Q_V(\boldsymbol{\alpha}, a)$ be the action value function

$$Q_V(\boldsymbol{\alpha}, a) = \boldsymbol{\alpha}^\top \mathbf{R}_a + \left(\boldsymbol{\beta}'_{a;\boldsymbol{\alpha}}\right)^\top \mathbf{V}(\boldsymbol{\alpha}'_{a,\mathcal{O}_0}). \quad (2)$$

Let a_V,a_W be actions satisfying $\hat{H}_n V = Q_V(\boldsymbol{\alpha}, a_V)$, $\hat{H}_n W = Q_W(\boldsymbol{\alpha}, a_W)$ for $\boldsymbol{\alpha} \in \mathcal{I}$. If $\boldsymbol{\beta'}_{a;\boldsymbol{\alpha}} \geq \mathbf{0}$ for all $a \in \mathcal{A}$, $V \leq W$ in $\mathcal{I}$ indicates

$$\begin{aligned}
&\Rightarrow \boldsymbol{\alpha}^\top \mathbf{R}_{a_V} + \left(\boldsymbol{\beta}'_{a_V;\boldsymbol{\alpha}}\right)^\top \mathbf{V}(\boldsymbol{\alpha}'_{a_V,\mathcal{O}_0}) \\
&\quad \leq \boldsymbol{\alpha}^\top \mathbf{R}_{a_V} + \left(\boldsymbol{\beta}'_{a_V;\boldsymbol{\alpha}}\right)^\top \mathbf{W}(\boldsymbol{\alpha}'_{a_V,\mathcal{O}_0}) \\
&\Rightarrow Q_V(\boldsymbol{\alpha}, a_V) \leq Q_W(\boldsymbol{\alpha}, a_V) \leq Q_W(\boldsymbol{\alpha}, a_W) \\
&\Rightarrow \hat{H}_n V(\boldsymbol{\alpha}) \leq \hat{H}_n W(\boldsymbol{\alpha}) \quad (3)
\end{aligned}$$

Since this holds for all $\boldsymbol{\alpha} \in \mathcal{I}$, $\hat{H}_n$ is isotonic in $\mathcal{I}$. □

The Bellman operator $\hat{H}_n^+$ is isotonic in the set of probability vectors $\mathcal{P} \subset \mathbb{R}^n$.

Theorem 2. *Suppose a value function $\hat{V}(\cdot)$ satisfies $\varepsilon = \sup_{\boldsymbol{\alpha}\in\mathcal{I}} \left|V^*(\boldsymbol{\alpha}) - \hat{V}(\boldsymbol{\alpha})\right|$. If $\boldsymbol{\beta'}_{a;\boldsymbol{\alpha}} \geq \mathbf{0}$ and $\max_{a\in\mathcal{A}} \left\|\boldsymbol{\beta'}_{a;\boldsymbol{\alpha}}\right\|_{L_1} \leq C$ for all $\boldsymbol{\alpha} \in \mathcal{I}$ and $a \in \mathcal{A}$, value iteration $\hat{H}_n\hat{V}$ has an error bound $\left|V^*(\boldsymbol{\alpha}) - (\hat{H}_n\hat{V})(\boldsymbol{\alpha})\right| \leq \gamma C\varepsilon$ for all $\boldsymbol{\alpha} \in \mathcal{I}$.*

Proof. $\left|V^*(\boldsymbol{\alpha}) - (\hat{H}_n\hat{V})(\boldsymbol{\alpha})\right|$ satisfies

$$\begin{aligned}
&\leq \max_{a\in\mathcal{A}} \left|Q^*(\boldsymbol{\alpha}, a) - \hat{Q}(\boldsymbol{\alpha}, a)\right| \\
&= \gamma \max_{a\in\mathcal{A}} \left|\boldsymbol{\beta}^\top_{a;\boldsymbol{\alpha}} \left(\mathbf{V}^*(\boldsymbol{\alpha}'_{a,\mathcal{O}_0}) - \hat{\mathbf{V}}(\boldsymbol{\alpha}'_{a,\mathcal{O}_0})\right)\right| \\
&\leq \gamma \max_{a\in\mathcal{A}} \boldsymbol{\beta}^\top_{a;\boldsymbol{\alpha}} \left|\mathbf{V}^*(\boldsymbol{\alpha}'_{a,\mathcal{O}_0}) - \hat{\mathbf{V}}(\boldsymbol{\alpha}'_{a,\mathcal{O}_0})\right| \\
&\leq \quad \gamma C\varepsilon
\end{aligned}$$

for all $\boldsymbol{\alpha} \in \mathcal{I}$. □

1-step value iteration using the corrected Bellman operator $\hat{H}_n^+$ has an error bound $\left|V^*(\boldsymbol{\alpha}) - (\hat{H}_n^+\hat{V})(\boldsymbol{\alpha})\right| \leq \gamma\varepsilon$ for all $\boldsymbol{\alpha} \in \mathcal{I}$ and d-step value iteration has $\left|V^*(\boldsymbol{\alpha}) - ((\hat{H}_n^+)^d\hat{V})(\boldsymbol{\alpha})\right| \leq \gamma^d\varepsilon$ for all $\boldsymbol{\alpha} \in \mathcal{I}$. $\hat{H}_n^+$ theoretically guarantees to run the kernel value iteration with finite horizons, though empirically $\hat{H}_n$ often worked.

References

[Porta. et al., 2006] J. M. Porta, N. Vlassis, and P. Poupart. Point-based value iteration for continuous POMDPs. *JMLR*, 7:2329–2367, 2006.