1 Proof of Theorem 1

We will need two lemmas for proving Theorem 1. The first one is Haussler’s bound given in \([1,\text{p. 103}](\text{Lemma 9, part (2)}).\)

**Lemma 1** (Haussler’s bound) Let \(Z_1,\ldots,Z_n\) be i.i.d random variables with range \(0 \leq Z_i \leq M\), \(E(Z_i) = \mu\), and \(\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} Z_i\), \(1 \leq i \leq n\). Assume \(\nu > 0\) and \(0 < \alpha < 1\). Then

\[
\Pr \left( d_{\nu}(\hat{\mu}, \mu) > \alpha \right) < 2e^{-\frac{\alpha^2}{2n/M}}
\]

where \(d_{\nu}(r,s) = \frac{|r-s|}{\nu + r + s}\). As a consequence,

\[
\Pr \left( \mu < \frac{1 - \alpha}{1 + \alpha} \hat{\mu} - \frac{\alpha}{1 + \nu} \right) < 2e^{-\frac{\alpha^2}{2n/M}}.
\]

Let \(\Pi_i\) be the class of policy trees in \(\Pi_{b_0,D,K}\) and having size \(i\). The next lemma bounds the size of \(\Pi_i\).

**Lemma 2** \(|\Pi_i| \leq i^{(i-2)}(|A||Z|)^i\).

**Proof.** Let \(\Pi_i'\) be the class of rooted ordered trees of size \(i\). \(|\Pi_i'|\) is not more than the number of all trees with \(i\) labeled nodes, because the in-order labeling of a tree in \(\Pi_i'\) corresponds to a labeled tree. By Cayley’s formula \([3]\), the number of trees with \(i\) labeled nodes is \(i^{(i-2)}\), thus \(|\Pi_i'| \leq i^{(i-2)}\). Recall the definition of a policy derivable from a DESPOT in Section 4 in the main text. A policy tree in \(\Pi_i\) is obtained from a tree in \(\Pi_i'\) by assigning the default policy to each leaf node, one of the \(|A|\) possible action labels to all other nodes, and one of at most \(|Z|\) possible labels to each edge. Therefore

\[
|\Pi_i| \leq i^{(i-2)} \cdot |A|^{i} \cdot |Z|^{(i-1)} \leq i^{(i-2)} (|A||Z|)^i.
\]

\(\square\)

In the following, we often abbreviate \(V_{\pi}(b_0)\) and \(\hat{V}_{\pi}(b_0)\) as \(V_{\pi}\) and \(\hat{V}_{\pi}\) respectively, since we will only consider the true and empirical values for a fixed but arbitrary \(b_0\). Our proof follows a line of reasoning similar to \([2]\).

**Theorem 1** For any \(\tau, \alpha \in (0,1)\) and any set \(\Phi_{b_0}\) of \(K\) randomly sampled scenarios for belief \(b_0\), every policy tree \(\pi \in \Pi_{b_0,D,K}\) satisfies

\[
V_{\pi}(b_0) \geq 1 - \frac{\alpha}{1 + \alpha} \hat{V}_{\pi}(b_0) - \frac{R_{\max}}{(1 + \alpha)(1 - \gamma)} \cdot \ln \left( \frac{4}{\tau} \right) + |\pi| \ln \left( KD |A| |Z| \right) \frac{\ln \left( KD |A| |Z| \right)}{\alpha K}.
\]

with probability at least \(1 - \tau\), where \(\hat{V}_{\pi}(b_0)\) denotes the estimated value of \(\pi\) under \(\Phi_{b_0}\).
Proof. Consider an arbitrary policy tree $\pi \in \Pi_{b_0,D,K}$. We know that for a random scenario $\phi$ for the belief $b_0$, executing the policy $\pi$ w.r.t. $\phi$ gives us a sequence of states and observations distributed according to the distributions $P(s'|s,a)$ and $P(z|s,a)$. Therefore, for $\pi$, its true value $V_{\pi}$ equals $\mathbb{E}(V_{\pi,\phi})$, where the expectation is over the distribution of scenarios. On the other hand, since $V_\pi = \frac{1}{K} \sum_{k=1}^{K} V_{\pi,\phi_k}$, and the scenarios $\phi_0, \phi_1, \ldots, \phi_K$ are independently sampled, Lemma 1 gives

$$\Pr \left( V_\pi < \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{\alpha}{1 + \alpha} \epsilon_{|\pi|} \right) < 2e^{-\alpha^2 \epsilon_{|\pi|} K/M} \quad (1)$$

where $M = R_{\max}/(1 - \gamma)$, and $\epsilon_i$ is chosen such that

$$2e^{-\alpha^2 \epsilon_{|\pi|} K/M} = \tau/(2\epsilon^2 |\Pi_i|). \quad (2)$$

By the union bound, we have

$$\Pr \left( \exists \pi \in \Pi_{b_0,D,K} \left[ V_\pi < \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{\alpha}{1 + \alpha} \epsilon_{|\pi|} \right] \right) \leq \sum_{i=1}^{\infty} \sum_{\pi \in \Pi_i} \Pr \left( V_\pi < \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{\alpha}{1 + \alpha} \epsilon_{|\pi|} \right).$$

By the choice of $\epsilon_i$'s and Inequality (1), the right hand side of the above inequality is bounded by $\sum_{i=1}^{\infty} |\Pi_i| \cdot [\tau/(2\epsilon^2 |\Pi_i|)] = \pi^2 \tau/12 < \tau$, where the well-known identity $\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$ is used. Hence,

$$\Pr \left( \exists \pi \in \Pi_{b_0,D,K} \left[ V_\pi < \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{\alpha}{1 + \alpha} \epsilon_{|\pi|} \right] \right) < \tau. \quad (3)$$

Equivalently, with probability $1 - \tau$, every $\pi \in \Pi_{b_0,D,K}$ satisfies

$$V_\pi \geq \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{\alpha}{1 + \alpha} \epsilon_{|\pi|}. \quad (4)$$

To complete the proof, we now give an upper bound on $\epsilon_{|\pi|}$. From Equation 2, we can solve for $\epsilon_{|\pi|}$ to get $\epsilon_i = \frac{R_{\max} \ln(4/\tau) + |\pi| \ln(KD|A||Z|)}{\alpha K}. \quad (5)$

Thus we have

$$\epsilon_{|\pi|} \leq \frac{R_{\max} \ln(4/\tau) + |\pi| \ln(KD|A||Z|)}{\alpha K}. \quad (6)$$

Combining this with Inequality (4), we get

$$V_\pi \geq \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{R_{\max} \ln(4/\tau) + |\pi| \ln(KD|A||Z|)}{(1 + \alpha)(1 - \gamma) \alpha K}. \quad (7)$$

This completes the proof. $\square$

2 Proof of Theorem 2

We need the following lemma for proving Theorem 2.

Lemma 3 For a fixed policy $\pi$ and any $\tau \in (0,1)$, with probability at least $1 - \tau$.

$$\hat{V}_\pi \geq V_\pi - \frac{R_{\max}}{1 - \gamma} \sqrt{\frac{2 \ln(1/\tau)}{K}}$$

Proof. Let $\pi$ be a policy and $V_\pi$ and $\hat{V}_\pi$ as mentioned. Hoeffding’s inequality gives us

$$\Pr \left( \hat{V}_\pi \geq V_\pi - \epsilon \right) \geq 1 - e^{-K\epsilon^2/(2M^2)}$$

This completes the proof. $\square$
Let \( \tau = e^{-K^2/(2M^2)} \) and solve for \( \epsilon \), then we get
\[
\Pr \left( \hat{V}_\pi \geq V_\pi - \frac{R_{\text{max}}}{1 - \gamma} \sqrt{\frac{2 \ln(1/\tau)}{K}} \right) \geq 1 - \tau.
\]
\( \square \)

**Theorem 2**  Let \( \pi^* \) be an optimal policy at a belief \( b_0 \). Let \( \pi \) be a policy derived from a DESPOT that has height \( D \) and are constructed from \( K \) randomly sampled scenarios for belief \( b_0 \). For any \( \tau, \alpha \in (0, 1) \), if \( \pi \) maximizes
\[
1 - \frac{\alpha}{1 + \alpha} \hat{V}(b_0) - \frac{R_{\text{max}}}{(1 + \alpha)(1 - \gamma)} \left( \frac{\ln(8/\tau) + |\pi| \ln(KD|A||Z|)}{\alpha K} \right),
\]
among all policies derived from the DESPOT, then
\[
V_\pi(b_0) \geq \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{R_{\text{max}}}{(1 + \alpha)(1 - \gamma)} \left( \frac{\ln(8/\tau) + |\pi| \ln(KD|A||Z|)}{\alpha K} \right).
\]
\[ \text{Proof.} \]
By Theorem 1, with probability at least \( 1 - \tau/2 \),
\[
V_\pi \geq \frac{1 - \alpha}{1 + \alpha} \hat{V}_\pi - \frac{R_{\text{max}}}{(1 + \alpha)(1 - \gamma)} \left( \frac{\ln(8/\tau) + |\pi| \ln(KD|A||Z|)}{\alpha K} \right).
\]

Suppose the above inequality holds on a random set of \( K \) scenarios. Note that there is a \( \pi' \in \Pi_{b_0,D,K} \) which is a subtree of \( \pi^* \) and has the same trajectories on these scenarios up to depth \( D \). By the choice of \( \pi \) in Inequality \( [5] \), it follows that with probability at least \( 1 - \tau/2 \),
\[
V_\pi \geq \frac{1 - \alpha}{1 + \alpha} \hat{V}_{\pi'} - \frac{R_{\text{max}}}{(1 + \alpha)(1 - \gamma)} \left( \frac{\ln(8/\tau) + |\pi'| \ln(KD|A||Z|)}{\alpha K} \right).
\]

Note that \( |\pi^*| \geq |\pi'| \), and \( \hat{V}_{\pi'} \geq \hat{V}_\pi - \gamma D R_{\text{max}}/(1 - \gamma) \) since \( \pi' \) and \( \pi^* \) only differ from depth \( D \) onwards, under the chosen scenarios. It follows that with probability at least \( 1 - \tau/2 \),
\[
V_\pi \geq \frac{1 - \alpha}{1 + \alpha} \left( \hat{V}_{\pi'} - \gamma D R_{\text{max}}/(1 - \gamma) \right) - \frac{R_{\text{max}}}{(1 + \alpha)(1 - \gamma)} \left( \frac{\ln(8/\tau) + |\pi^*| \ln(KD|A||Z|)}{\alpha K} \right).
\]
\[ \text{By Lemma 3} \] with probability at least \( 1 - \tau/2 \), we have
\[
\hat{V}_{\pi'} \geq V_\pi - \frac{R_{\text{max}}}{1 - \gamma} \sqrt{\frac{2 \ln(2/\tau)}{K}}.
\]
By the union bound, with probability at least \( 1 - \tau \), both Inequality \( [7] \) and Inequality \( [8] \) hold, which imply Inequality \( [6] \) holds. This completes the proof. \( \square \)

**References**


