6 Appendix

In the appendix section, we detail out the proof of our main theorem and lemma. We also show a comparison on algorithm 1 and algorithm 2 to showcase the benefits of applying matrix sensing technique.

6.1 Proof for Lemma 3.1

Lemma 6.1. *Given a POMDP $\psi = \langle \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{O}}, \mathbf{r}, \mathcal{A}, \mathcal{O}, S, \boldsymbol{\mu}, \gamma \rangle$ of size k and a sampling policy Π induced by $\boldsymbol{\Pi} \in [0,1]^{\mathcal{S}\times\mathcal{A}}$, there exists a WFA $B = \langle \boldsymbol{\beta}, \{\mathbf{B}_\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau} \rangle$ with k states that realizes the function $g(h) = \tilde{R}(h)\mathbb{P}^{\Pi}(h)$, where $\Sigma = \mathcal{A}\times\mathcal{O}$ and $h \in \Sigma^*$.*

Proof. Let s^i denote the i^{th} state and let

$$\tilde{\mathbf{O}}_{ao} = \text{diag}(\boldsymbol{\mathcal{O}}_{s^1,a,o}, \boldsymbol{\mathcal{O}}_{s^2,a,o}, \cdots, \boldsymbol{\mathcal{O}}_{s^k,a,o}),$$

$$\tilde{\mathbf{M}}_a = \text{diag}(\boldsymbol{\Pi}_{s^1,a}, \boldsymbol{\Pi}_{s^2,a}, \cdots, \boldsymbol{\Pi}_{s^k,a}).$$

We can construct a WFA $B = \langle \boldsymbol{\beta}^\top, \{\mathbf{B}_\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau} \rangle$ such that: $\boldsymbol{\beta}^\top = \boldsymbol{\mu}^\top$, $\mathbf{B}_\sigma = \mathbf{B}_{ao} = \tilde{\mathbf{M}}_a \boldsymbol{\mathcal{T}}_{:,a,:} \tilde{\mathbf{O}}_{ao}$, $\boldsymbol{\tau} = \mathbf{r}$. Then by construction, one can check that the WFA B computes the function g, which also shows that the rank of the function g is at most k. □

6.2 Proof for Theorem 3.2

Theorem 6.2. *Given a POMDP ψ of size k, a sampling policy Π and a WFA $B = \langle \boldsymbol{\beta}^\top, \{\mathbf{B}_\sigma\}_{\sigma\in\Sigma}, \boldsymbol{\tau} \rangle$ realizing the function $g : h \mapsto \tilde{R}(h)\mathbb{P}^{\Pi}(h)$ such that the spectral radius $\rho(\gamma \sum_{\sigma\in\Sigma} \mathbf{B}_\sigma) < 1$, the WFA $A = \langle \boldsymbol{\beta}^\top, \{\mathbf{B}_\sigma\}_{\sigma\in\Sigma}, (\mathbf{I} - \gamma \sum_{\sigma\in\Sigma} \mathbf{B}_\sigma)^{-1}\boldsymbol{\tau} \rangle$ of size k realizes the function $\tilde{V}^{\Pi}(h) = \sum_{z\in\Sigma^*} \gamma^{|z|} \tilde{R}(hz)\mathbb{P}^{\Pi}(hz)$.*

Proof. By definition of the function $\tilde{V}^{\Pi}$, we have:

$$\begin{aligned}
\tilde{V}^{\Pi}(h) &= \sum_{z\in\Sigma^*} \gamma^{|z|} \tilde{R}(hz)\mathbb{P}^{\Pi}(hz) \\
&= \sum_{z\in\Sigma^*} \gamma^{|z|} \boldsymbol{\beta}^\top \mathbf{B}_h \mathbf{B}_z \boldsymbol{\tau} \\
&= \boldsymbol{\beta}^\top \mathbf{B}_h \Big(\sum_{z\in\Sigma^*} \gamma^{|z|} \mathbf{B}_z \Big) \boldsymbol{\tau} \\
&= \boldsymbol{\beta}^\top \mathbf{B}_h \Big(\sum_{i=0}^{\infty} (\gamma \sum_{\sigma\in\Sigma} \mathbf{B}_\sigma)^i \Big) \boldsymbol{\tau} \\
&= \boldsymbol{\beta}^\top \mathbf{B}_h (\mathbf{I} - \gamma \sum_{\sigma\in\Sigma} \mathbf{B}_\sigma)^{-1} \boldsymbol{\tau}
\end{aligned}$$

Here we applied Neumann identity: $\sum_{i=0}^{\infty} \mathbf{T}^i = (\mathbf{I} - \mathbf{T})^{-1}$, which holds when $\rho(\mathbf{T}) < 1$. Therefore, the WFA $A = \langle \boldsymbol{\beta}^\top, \{\mathbf{B}_\sigma\}_{\sigma\in\Sigma}, (\mathbf{I} - \gamma \sum_{\sigma\in\Sigma} \mathbf{B}_\sigma)^{-1}\boldsymbol{\tau} \rangle$ realizes the function $\tilde{V}^{\Pi}$. □

6.3 Time complexity analysis

The need for a compressed representation is particularly evident in large data settings, that is when the Hankel matrix becomes too big and sparse to handle without compression. Figure 3 compares the uncompressed UQF algorithm (algorithm 1) with the compressed version (algorithm 2). As the figure shows, the compressed representation boosts the computation time significantly with almost no cost to the returns collected by the policy.

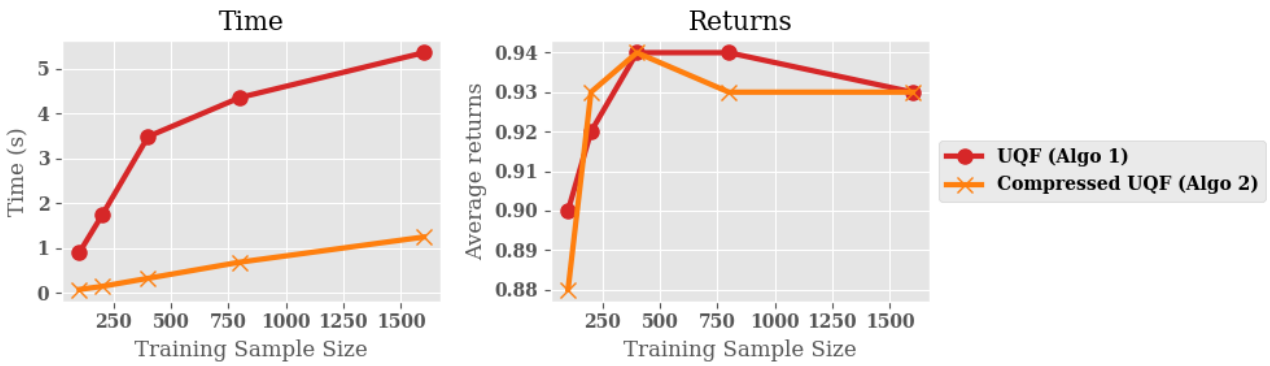

Figure 3: Comparison between Algorithm 1 and Algorithm 2 in terms of computation time and performance (average returns).