Appendix

Lemma 5 (Davis Kahn $\sin\Theta$ bound [21]) *Let $\Sigma, \hat{\Sigma} \in \mathbb{R}^{D\times D}$ be symmetric, with eigen values $\lambda_1 \geq \cdots \lambda_D$ and $\hat{\lambda}_1 \geq \cdots \hat{\lambda}_D$, where $\Sigma v_i = \lambda_i v_i$, $\forall i \in [D]$ and $\hat{\Sigma}\hat{v}_i = \hat{\lambda}_i \hat{v}_i$, $\forall i \in [D]$. Further, define $E = [v_1, \cdots, v_d] \in \mathbb{R}^{D\times d}$, $\hat{E} = [\hat{v}_1, \cdots, \hat{v}_d]$ and assume that $\delta_* = \lambda_d - \lambda_{d+1} > 0$.*

Then, we have that,

$$\|\sin\Theta(E, \hat{E})\|_F \leq \frac{2\sqrt{d}}{\delta_*}\|\Sigma - \hat{\Sigma}\|_2 \tag{13}$$

Lemma 6 (Concentration of sub-Gaussian covariance) *Let $\mathcal{D}$ be a zero-mean sub-Gaussian distribution in $\mathbb{R}^D$ s.t. for $X \sim \mathcal{D}$, we have that, $\|\langle X, x\rangle\|_{\psi_2} \leq K\|\langle X, x\rangle\|_2$, $\forall x \in \mathbb{R}^D$. Denote the covariance matrix $\Sigma = \mathbb{E}_{X\sim\mathcal{D}}\left[XX^T\right]$ and $\Sigma_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$, where $X_i \overset{i.i.d}{\sim} \mathcal{D}$. Then $\forall \delta > 0$, we have that,*

$$\|\Sigma_n - \Sigma\|_2 \leq CK^2\|\Sigma\|\left(\sqrt{\frac{D + \log\frac{2}{\delta}}{n}} + \frac{D + \log\frac{2}{\delta}}{n}\right)$$

with probability $\geq 1 - \delta$, for some fixed small constant C.

Proof: Refer to Theorem 4.7.1 and Exercise 4.7.3 in Vershynin [17]. □

Note that in large data regime, i.e. $n >> D$, only the first term dominates while in the small data and large dimension regime the second terms dominates. In large data regime, for any given error parameter $\epsilon > 0$, one can sample n large enough and so the first term dominates the second one.

Lemma 7 *(Concentration of Projection matrix) Let $X \in \mathbb{R}^D$ be a zero mean sub-gaussian random vector with bounded sub-gaussian norm i.e.*

$$\exists K \in \mathbb{R} \text{ s.t. } \|\langle X, v\rangle\|_{\psi_2} \leq K\|\langle X, v\rangle\|_{L^2}\ \forall v \text{ s.t. } \|v\|_2 = 1$$

Given n i.i.d sample $\{X_i\}|_{i=1}^n \overset{i.i.d}{\sim} P_X$, let $\Sigma_n = \frac{1}{n}\sum_{i=1}^n X_i X_i^\top$ be the empirical estimate of the covariance matrix $\Sigma = \mathbb{E}[XX^\top]$. Further, let the eigen decomposition of Σ, Σ_n be given as,

$$\Sigma = \sum_{i=1}^D \lambda_i u_i u_i^\top, \qquad \Sigma_n = \sum_{i=1}^D \hat{\lambda}_i \hat{u}_i \hat{u}_i^\top$$

and $\mathit{Proj}_E = \sum_{i=1}^d u_i u_i^\top$ and $\mathit{Proj}_{E_n} = \sum_{i=1}^d \hat{u}_i \hat{u}_i^\top$ denote the projection operator onto the top d true and empirical eigen subspaces $E = span(\{u_i\}|_{i=1}^d)$ and $E_n = span(\{\hat{u}_i\}|_{i=1}^d)$ respectively. Then, if $n \geq \frac{CdK^4\|\Sigma\|_2^2}{{\delta_*}^2\cdot\epsilon^2}\left(D + \log\frac{2}{\delta}\right)$, we have,

$$w.p \geq 1 - \delta, \quad \|\mathit{Proj}_E - \mathit{Proj}_{E_n}\|_2 \leq \epsilon \tag{14}$$

where δ_* is the eigen gap between top d and the rest eigen subspace of true covariance matrix Σ.

Proof: Let $E^\top E_n = U\cos\Theta(E, E_n)V^T$ denote the singular value decomposition of $E^\top E_n$ and $\cos\Theta(E, E_n) = \mathrm{diag}(\cos\theta_1, \cdots, \cos\theta_d)$ be the diagonal matrix with the cosine of principal angles $\{\theta_1, \cdots, \theta_d\}$ between E and E_n subspaces as its diagonal entries. With this definition, it is easy to show that,

$$\|\mathit{Proj}_E - \mathit{Proj}_{E_n}\|_2 = \sqrt{2}\|\sin\Theta(E, E_n)\|_2$$

Next, using $\sin\Theta$ variant of Davis Kahn theorem 5, we have,

$$\|\sin\Theta(E, E_n)\|_F \leq \frac{2\sqrt{d}}{\delta_*}\|\Sigma_n - \Sigma\|_2 \tag{15}$$

$$\implies \|\mathit{Proj}_E - \mathit{Proj}_{E_n}\|_2 \leq \sqrt{2}\|\sin\Theta(E, E_n)\|_F \leq \frac{2\sqrt{2d}}{\delta_*}\|\Sigma_n - \Sigma\|_2 \tag{16}$$

Note that we have used the fact that L_2 norm is upper bounded by Frobenious norm. Next, using concentration bound for covariance matrix of sub-gaussian random vectors 6(in large sample regime), we have, $w.p. \geq 1-\delta$,

$$\|\Sigma_n - \Sigma\|_2 \leq CK^2 \|\Sigma\|_2 \sqrt{\frac{D + \log(\frac{2}{\delta})}{n}} \tag{17}$$

Using it along with (16), gives us, $w.p. \geq 1-\delta$,

$$\|\mathit{Proj}_E - \mathit{Proj}_{E_n}\|_2 \leq \frac{2\sqrt{2d}}{\delta_*} CK^2 \|\Sigma\|_2 \sqrt{\frac{D + \log(\frac{2}{\delta})}{n}} \tag{18}$$

Finally, choosing $n \geq \frac{CdK^4 \|\Sigma\|^2}{\delta_*{}^2 \cdot \epsilon^2} \left(D + \log \frac{2}{\delta}\right)$, establishes the lemma.

□

Lemma 8 (Performance difference lemma for a general policy) *Let π, π' be two time and history dependent non-Markovian policies acting in environment MDP $\mathcal{M}$. Then, the performance difference between the value function of these policy in environment $\mathcal{M}$ is given by :*

$$V^{\pi}_{\mathcal{M}}(s_0) - V^{\pi'}_{\mathcal{M}}(s_0) = \sum_{t=0}^{\infty} \gamma^t \mathop{\mathbb{E}}_{s_{0:t} \sim d^{\pi, 0:t}_{\mathcal{M}}} \left[Q^{\pi'}_{\mathcal{M}}(s_{0:t}, \pi(s_{0:t})) - Q^{\pi'}_{\mathcal{M}}(s_{0:t}, \pi'(s_{0:t})) \right] \tag{19}$$

Further, if π, π' are Markovian, we have,

$$V^{\pi}_{\mathcal{M}}(s_0) - V^{\pi'}_{\mathcal{M}}(s_0) = \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi}_{\mathcal{M}}} \left[Q^{\pi'}_{\mathcal{M}}(s_t, \pi(s_t)) - V^{\pi'}_{\mathcal{M}}(s_t) \right] \tag{20}$$

Proof: We provide a complete proof of the performance difference lemma for self sufficiency.

$$\begin{aligned}
V^{\pi}_{\mathcal{M}}(s_0) - V^{\pi'}_{\mathcal{M}}(s_0) &= \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi \circ f(s_t))\right] - V^{\pi' \circ f}_{\mathcal{M}}(s_0) && (21)\\
&= \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left(R(s_t, \pi \circ f(s_{0:t})) + V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t}) - V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t})\right)\right] - V^{\pi' \circ f}_{\mathcal{M}}(s_0) && (22)\\
&= \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left(R(s_t, \pi \circ f(s_{0:t})) + \gamma V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t+1}) - V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t})\right)\right] && (23)\\
&= \sum_{t=0}^{\infty} \gamma^t \mathbb{E}\left[\left(R(s_t, \pi \circ f(s_{0:t})) + \gamma V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t+1}) - V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t})\right)\right] && (24)\\
&= \sum_{t=0}^{\infty} \gamma^t \mathbb{E}\left[\left(R(s_t, \pi \circ f(s_{0:t})) + \gamma V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t+1}) - V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t})\right)\right] && (25)\\
&= \sum_{t=0}^{\infty} \gamma^t \mathop{\mathbb{E}}_{s_{0:t} \sim d^{\pi \circ f, 0:t}_{\mathcal{M}}} \left[\mathop{\mathbb{E}}_{s_t+1} \left[\left(R(s_t, \pi \circ f(s_{0:t})) + \gamma V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t+1}) - V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t})\right) \middle| s_{0:t}\right]\right] && (26)\\
&= \sum_{t=0}^{\infty} \gamma^t \mathop{\mathbb{E}}_{s_{0:t} \sim d^{\pi \circ f, 0:t}_{\mathcal{M}}} \left[Q^{\pi' \circ f}_{\mathcal{M}}(s_{0:t}, \pi \circ f(s_{0:t})) - V^{\pi' \circ f}_{\mathcal{M}}(s_{0:t}) \right] && (27)
\end{aligned}$$

Further, if π, π' are Markovian, it simplifies to

$$\begin{aligned}
V^{\pi}_{\mathcal{M}}(s_0) - V^{\pi'}_{\mathcal{M}}(s_0) &= \sum_{t=0}^{\infty} \gamma^t \mathop{\mathbb{E}}_{s \sim d^{\pi, t}_{\mathcal{M}}} \left[Q^{\pi'}_{\mathcal{M}}(s_t, \pi(s_t)) - V^{\pi'}_{\mathcal{M}}(s_t) \right] && (28)\\
&= \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi}_{\mathcal{M}}} \left[Q^{\pi'}_{\mathcal{M}}(s_t, \pi(s_t)) - V^{\pi'}_{\mathcal{M}}(s_t) \right] && (29)
\end{aligned}$$

□