Recap of FQE: $Q^{\pi} \to J(\pi) = \mathbb{E}_{d_0}[Q^{\pi}(s,\pi)]$

① $\|f_k - \mathcal{T}^{\pi} f_{k-1}\|_{2,\mu} \le O\left(\frac{1}{n^{1/4}}\right) \to \frac{1}{\sqrt{n}}$

$$\|f_k - \mathcal{T}^{\pi} f_{k-1}\|_{2,\mu}^2 = O\left(\frac{1}{\sqrt{n}}\right) \to \frac{1}{n}$$

② $\|\cdot\|_{p,\mu} = \left(\mathbb{E}_{\mu}[|\cdot|^p]\right)^{1/p}$

$$\|\cdot\|_{1,\mu} \le \|\cdot\|_{2,\mu} \le \|\cdot\|_{\infty}$$

$$\text{AM} \le \text{QM} \le \max$$

$$\|\cdot\|_{p,\mu} \le \left(\left\|\frac{\nu}{\mu}\right\|_{\infty}\right)^{1/p} \|\cdot\|_{p,\nu}$$

(3) $\mathbb{E}_{d_t^{\pi}}[f_k - \mathcal{T}^{\pi} f_{k-1}]$

$$\le \|f_k - \mathcal{T}^{\pi} f_{k-1}\|_{2,d_t^{\pi}}$$

$$\le \sqrt{\left\|\frac{d_t^{\pi}}{\mu}\right\|_{\infty}} \|f_k - \mathcal{T}^{\pi} f_{k-1}\|_{2,\mu}$$

FQI: $f_k \leftarrow \underset{f \in \mathcal{F}}{\operatorname{argmin}} \mathbb{E}_\mu\left[\left(f(s,a) - r - \gamma \max_{a'} f_{k-1}(s',a')\right)^2\right]$

$\pi_{f_k}(s) = \underset{a}{\operatorname{argmax}}\ f_k(s,a).$

Output: $\hat{\pi} = \pi_{f_K} \circ \pi_{f_{K-1}} \circ \pi_{f_{K-2}} \circ \cdots \circ \pi_{f_0} \circ$ anything.

Goal: $J(\pi^*) - J(\hat{\pi})$. d_0.

Analysis: $\|f_k - \mathcal{T} f_{k-1}\|_{2,\mu} \le \varepsilon.\ \forall k.$

Given an initial distribution $d_0 \in \Delta(\mathcal{S})$, let $J(\pi) := \mathbb{E}[\sum_{t=1}^{\infty} \gamma^{t-1} r_t | \pi, s_1 \sim d_0]$. (Note that this definition applies to non-stationary π.) **Show that** for any (possibly non-stationary) π,

$$J(\pi) - J(\widehat{\pi}) \le \sum_{t=1}^{K} \gamma^{t-1} \left(\mathbb{E}_{d_t^{\pi}}[\mathcal{T} f_{K-t} - f_{K-t+1}] + \mathbb{E}_{d_t^{\widehat{\pi}}}[f_{K-t+1} - \mathcal{T} f_{K-t}] \right) + \gamma^K V_{\max}. \tag{3}$$

π^* $\quad d_t^{\pi}$ $\quad d_t^{\hat{\pi}}$

(let $\pi = \pi^*$).

$$\le \sum_{t=1}^{K} \gamma^{t-1} \Big(\|\mathcal{T} f_{K-t} - f_{K-t+1}\|_{2, d_t^{\pi^*}} + \|\mathcal{T} f_{K-t} - f_{K-t+1}\|_{2, d_t^{\hat{\pi}}} \Big)$$

$\to \|\cdot\|_{2,\mu}.$

$$\le \sum_{t=1}^{K} \gamma^{t-1} \left(\sqrt{\left\|\frac{d_t^{\pi^*}}{\mu}\right\|_\infty} \varepsilon + \sqrt{\left\|\frac{d_t^{\hat{\pi}}}{\mu}\right\|_\infty} \varepsilon \right).$$

$$\left\|\frac{d_t^{\hat{\pi}}}{\mu}\right\|_\infty \le \max_{\pi \in \Pi} \left\|\frac{d_t^{\pi}}{\mu}\right\|_\infty .$$

$$J(\pi^*) - J(\pi_{f_k}).$$

$$= \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_{f_k}}} \left[V^*(s) - \underbrace{Q^*(s, \pi_{f_k})}_{\le} \right]$$

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, \begin{array}{l} s_1 = s \\ a_1 \sim \pi_{f_k} \\ a_{2:\infty} \sim \pi^* \end{array}\right]$$

$$\le \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_{f_k}}} \left[V^*(s) - V^{\hat{\pi}}(s) \right].$$

$\mathcal{T} f \in \mathcal{F} \quad \forall f \in \mathcal{F}. \quad \not\Rightarrow$ FQI converge.

$$f_k = f_{k-1}. \qquad \cancel{\|f_k - \mathcal{T} f_k\| \le \varepsilon}$$

$$f_1 \; f_2 \; f_3 \; \cdots \; f_{k-1} = f_k \longrightarrow f \; f \; f \; f$$

$$\min_f \left(\mathbb{E}\left[\left(f(s,a) - r - \gamma \max_{a'} f(s',a')\right)^2\right] - \min_{g \in \mathcal{G}} \left[\mathbb{E}\left(g - \dots\right)^2\right]\right)$$

$\mathcal{T}^\pi f \in \mathcal{G}$.

$$\|\hat{f} - \mathcal{T}\hat{f}\|_{2,\mu} \le \varepsilon.$$

---

① $\mathcal{T}f \in \mathcal{F}\ \forall f \in \mathcal{F}$ (or $\mathcal{T}^\pi f \in \mathcal{F}\ \forall f \in \mathcal{F}$).

② $\max_{\pi, t} \left\|\frac{d_t^\pi}{\mu}\right\|_\infty \le C$

$Q^* \in \mathcal{F}$?

$\left\|\frac{d^{\pi^*}}{\mu}\right\|_\infty \le C.$