Convergence of tabular methods

Let's consider the following undiscounted ∞-horizon problem:

$$\max_{\rho} \mathbb{E}\left[\sum_{k=0}^{\infty} r(X_k, U_k)\right]$$

$$\text{s.t.} \quad X_{k+1} \sim f(x_{k+1} \mid x_k, u_k)$$

$$U_k = \rho(X_k)$$

Let's assume $r \leq 0$

Let's do tabular VI:

$$\rho_i(x) := \arg\max_u \left\{ r(x,u) + \mathbb{E}\left[V_i(X_+^u)\right] \right\} \quad \forall x \in \mathcal{X}$$

$$V_{i+1}(x) := r(x, \rho_i(x)) + \mathbb{E}\left[V_i\left(X_+^{\rho_i(x)}\right)\right]$$

Reminder: $X_+^u \sim f(x_+ \mid x, u)$

Let's consider $\eta : \mathcal{X} \to \mathcal{U}$, an arbitrary policy

Its value reads:

$$V^{\eta}(x) = \mathbb{E}_f\left[\sum_{k=0}^{\infty} r(X_k, \eta(X_k)) \,\middle|\, X_0 = x\right]$$

Since V^{η} may fail to converge, we consider so-called

admissible policies.

Admissible $\eta \Rightarrow V^{\eta} > -\infty$

Let's consider generic iterations of the kind:

$$\Lambda_{i+1}(x) := r(x, \eta(x)) + \mathbb{E}\left[\Lambda_i\left(X_i^{\eta(x)}\right)\right]$$

Start with $\Lambda_0 \equiv 0$

Let's benchmark our VI scheme using Λ_i.

Iteration $i = 1$:

$$V_1(x) = r(x, \rho_0(x))$$

let's say ρ_0 was chosen so as to satisfy $\forall x \in \mathcal{X}^- \quad r(x, \rho_0(x)) \geq r(x, \eta(x))$

At the same time,

$$\Lambda_1(x) = r(x, \eta(x))$$

$$\Rightarrow \quad V_1 \geq \Lambda_1$$

Iteration 2:

$$V_2(x) = r(x, \rho_1(x)) + \mathbb{E}\left[V_1\left(X_+^{\rho_1(x)}\right)\right]$$

$$= r(x, \rho_1(x)) + \mathbb{E}\left[r\left(X_+^{\rho_1(x)}, \rho_0\left(X_+^{\rho_1(x)}\right)\right)\right]$$

$$\rho_1(x) = \arg\max_u \left\{ r(x,u) + \mathbb{E}\left[r\left(X_+^u, \rho_0(X_+^u)\right)\right] \right\}$$

So since ρ_1 is a maximizer,

$$r(x, \rho_1(x)) + \mathbb{E}\left[r\left(X_+^{\rho_0(x)}, \rho_1\left(X_+^{\rho_0(x)}\right)\right)\right] \geq$$

$$r(x, \eta(x)) + \mathbb{E}\left[r\left(X_+^{\rho_0(x)}, \eta\left(X_+^{\rho_0(x)}\right)\right)\right] \geq$$

$$r(x, \eta(x)) + \mathbb{E}\left[r\left(X_+^{\eta(x)}, \eta\left(X_+^{\eta(x)}\right)\right)\right]$$

So, we may conclude that

$$V_2 \geq \Lambda_2$$

Let's assume $V_i \geq \Lambda_i$

Consider

$$\rho_i(x) = \arg\max_u \left\{ r(x,u) + \mathbb{E}\left[V_i(X_+^u) \right] \right\}$$

$$V_{i+1}(x) = r(x, \rho_i(x)) + \mathbb{E}\left[V_i\left(X_+^{\rho_i(x)}\right) \right]$$

$$\Lambda_{i+1}(x) = r(x, \eta(x)) + \mathbb{E}\left[\Lambda_i\left(X_+^{\eta(x)}\right) \right]$$

By optimality of ρ_i,

$$r(x, \rho_i(x)) + \mathbb{E}\left[V_i\left(X_+^{\rho_i(x)}\right) \right] \geq$$

$$r(x, \eta(x)) + \mathbb{E}\left[V_i\left(X_+^{\eta(x)}\right) \right]$$

By the induction hypothesis,

$$r(x, \eta(x)) + \mathbb{E}\left[V_i\left(X_+^{\eta(x)}\right) \right] \geq$$

$$r(x, \eta(x)) + \mathbb{E}\left[\Lambda_i\left(X_+^{\eta(x)}\right) \right]$$

Which means,

$$V_{i+1}(x) \geq \Lambda_{i+1}(x), \quad \forall x \in \mathbb{X}$$

So, a proof by induction is complete.

Since this holds for any η (admissible), it also holds for the optimal policy ρ^*!

The benchmark iterations read:

$$\Lambda_{i+1}(x) = r(x, \rho^*(x)) + \mathbb{E}\left[\Lambda_i\left(X_+^{\rho^*(x)}\right)\right]$$

Unwrapping the recursion yields:

$$\Lambda_{i+1}(x) = \mathbb{E}\left[\sum_{k=0}^{i-1} r(X_k, \rho^*(X_k)) \,\middle|\, X_0 = x\right]$$

Adding the infinite tail, and keeping in mind that $r \leq 0$,

$$\Lambda_i(x) \geq \mathbb{E}\left[\sum_{k=0}^{\infty} r(X_k, \rho^*(X_k)) \,\middle|\, X_0 = x\right]$$

$$= V^*(x)$$

Looking at two successive iterations:

$$V_{i+1}(x) = r(x, \rho_i(x)) + \mathbb{E}\left[V_i\left(X_+^{\rho_i(x)}\right)\right]$$

$$V_i(x) = r(x, \rho_{i-1}(x)) + \mathbb{E}\left[V_{i-1}\left(X_+^{\rho_{i-1}(x)}\right)\right]$$

Remember, ρ_{i-1} was optimal on its iteration:

$$\rho_{i-1}(x) = \underset{u}{\operatorname{argmax}} \left\{ r(x,u) + \mathbb{E}\left[V_{i-2}\left(X_+^{u}\right)\right]\right\}$$

We thus have

$$r(x, \rho_{i-1}(x)) + \mathbb{E}\left[V_{i-2}\left(X_+^{\rho_{i-1}(x)}\right)\right] \geq$$

$$r\left(x, \rho_i(x)\right) + \mathbb{E}\left[V_{i-1}\left(X_+^{\rho_i(x)}\right)\right]$$

So,

$$V_i(x) \geq r(x, \rho_i(x)) + \mathbb{E}\left[V_{i-1}\left(X_+^{\rho_i(x)}\right)\right]$$

Let's do some more unwrapping:

$$V_i(x) = \mathbb{E}\left[\sum_{k=0}^{i-1} r\left(X_k, \rho_{i-1-k}(X_k)\right) \middle| X_0 = x\right]$$

$$V_{i-1}(x) = \mathbb{E}\left[\sum_{k=0}^{i-2} r\left(X_k, \rho_{i-2-k}(X_k)\right) \middle| X_0 = x\right]$$

Now shift the starting state to X_+^u:

$$V_i(X_+^u) = \mathbb{E}\left[\sum_{k=1}^{i} r\left(X_k, \rho_{i-k}(X_k)\right) \middle| X_0 = X_+^u\right]$$

$$V_{i-1}(X_+^u) = \mathbb{E}\left[\sum_{k=1}^{i-1} r\left(X_k, \rho_{i-1-k}(X_k)\right) \middle| X_0 = X_+^u\right]$$

Add the ith reward to the last expression to get:

$$\mathbb{E}\left[\sum_{k=1}^{i-1} r\left(X_k, \rho_{i-1-k}(X_k)\right) \middle| X_0 = X_+^u\right] \geq$$

$$\mathbb{E}\left[\sum_{k=1}^{i-1} r\left(X_k, \rho_{i-1-k}(X_k)\right) + r\left(X_i, \rho_{i-1}(X_i)\right) \middle| X_0 = X_+^u\right] =$$

$$\mathbb{E}\left[\sum_{k=1}^{i} r\left(X_k, \rho_{i-k}(X_k)\right) \middle| X_0 = X_+^u\right] =$$

$$V_i(X_+^u)$$

Therefore,

$$V_i(x) \geq r(x, \rho_i(x)) + \mathbb{E}\left[V_{i+1}\left(X_+^{\rho_i(x)}\right)\right]$$

$$\geq r(x, \rho_i(x)) + \mathbb{E}\left[V_i\left(X_+^{\rho_i(x)}\right)\right]$$

$$= V_{i+1}(x)$$

So, we have:

$$\forall i,\ x \in \mathcal{X} \quad V_i(x) \geq V_{i+1}(x)$$

$$V_i(x) \geq V^*(x) > -\infty$$

To show convergence, apply the monotone convergence theorem

So, $V_i \to V_\infty$

We know $V_\infty \geq V^*(x)$

Since V^* is the optimal value function, it has to be the greatest, i.e., $V^*(x) \geq V_\infty(x), \forall x \in \mathcal{X}$,

So $V_\infty = V^*$!