Convergence of VI

V_i is pointwise increasing

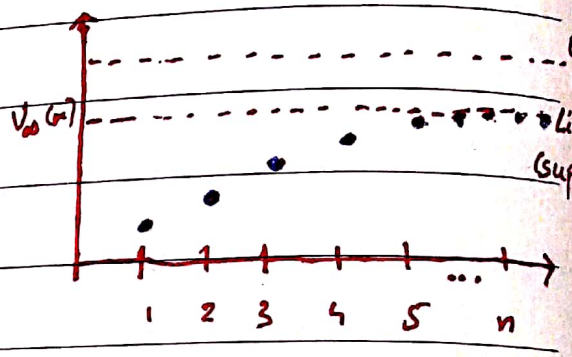
&

Bounded

⇓

pointwise convergent

Monotone convergence Theorem



$\forall i, x$

$$V_i(x) \le V_{i+1}(x)$$

$$V_i(x) \le V(x) < \infty$$

$$\downarrow$$

$$V_\infty(x) = V(x) \qquad \forall x$$

Moreover

Policy: $\mu_\infty = \mu^*$

∞-horizon undiscounted optimal control problem

$$\min_{\mu} \mathbb{E}\left[\sum_{k=0}^{\infty} \rho(X_k, U_k)\right], \quad \rho(x,u) \geq 0 \;\; \forall x,u$$

$$X_+ \sim P_X(X_+ | x, u)$$

$$U_+ \sim P_U^{\mu(x)}(u|x)$$

parameters of policy model '$\mu(x)$' depend upon state, x.

$$\mu : \mathcal{X} \to \Pi$$

space of policy PDF parameters

VI algorithm: $\forall x$

$$\mu_i(x) := \arg\min_{\theta} \left\{ \mathbb{E}_{U \sim P_U^{\theta}} \left[\rho(x,U) + V_i(X_+^U) \right] \right\}$$

$$V_{i+1} := \mathbb{E}_{U \sim P_U^{\mu_i(x)}} \left[\rho(x,U) + V_i(X_+^U) \right]$$

η arbitrary, "admissible" policy

$$J(x|\eta) = \mathbb{E}_{U_k \sim p_U^{\eta(x_k)}} \left[\sum_{k=0}^{\infty} \rho(x_k, U_k) \,\middle|\, x_0 = x \right]$$

should be finite !

→ What properties comprise an admissible policy ?

Comparing iteration by iteration

$$V_{i+1} := \mathbb{E}_{U \sim p_U^{\mu_i(x)}} \left[\rho(x, U) + V_i(x^+) \right]$$

$$\Lambda_{i+1} := \mathbb{E}_{U \sim p_U^{\eta(x)}} \left[\rho(x, U) + \Lambda_i(x^+) \right]$$

$$\Lambda_0 = 0$$

① $$V_1(x) = \mathbb{E}_{U \sim p_U^{\mu_0(x)}} \left[\rho(x, U) \right]$$

① $$\Lambda_1(x) = \mathbb{E}_{U \sim p_U^{\eta(x)}} \left[\rho(x, U) \right]$$

where

$$\mu_0(x) = \operatorname*{argmin}_{\sigma} \mathbb{E}_{U \sim p_U^{\sigma}} \left[\rho(x, U) \right]$$

Because of minimizer (or optimality)

$\mu_0(x)$

$$V_1(x) \leq \Lambda_1(x)$$

(2)

$$V_2(x) = \mathbb{E}_{U \sim \rho_U^{\mu_1(x)}} \left[\rho(x, U) + V_1(x_+^U) \right]$$

$$= \mathbb{E}_{U \sim \rho_U^{\mu_1(x)}} \left[\rho(x, U) + \mathbb{E}_{U \sim \rho_U^{\mu_0(x_+^U)}} \left[\rho(x_+^+, U) \mid X = x \right] \right]$$

$$= \mathbb{E}_{\substack{U \sim \rho_U^{\mu_1(x)} \\ U_+ \sim \rho_U^{\mu_0(x_+^U)}}} \left[\rho(x, U) + \rho(x_+^U, U_+) \right]$$

$$\mu_1(x) = \operatorname*{argmin}_{\sigma} \mathbb{E}_{\substack{U \sim \rho_U^{\sigma(x)} \\ U_+ \rho_U^{\mu_0(x_+^U)}}} \left[\rho(x, U) + \rho(x_+^U, U_+) \right]$$

Due to optimality of $\mu_1(x)$

$$\mathbb{E}_{\substack{U \sim \rho_U^{\mu_1(x)} \\ U_+ \sim \rho_U^{\mu_0(x_+^U)}}} \left[\rho(x, U) + \rho(x_+^U, U_+) \right] \leq \mathbb{E}_{\substack{U \sim \rho_U^{\eta(x)} \\ U_+ \sim \rho_U^{\mu_0(x_+^U)}}} \left[\rho(x, U) + \rho(x \right.$$

can be cho

$U_+ \sim \rho_U^{\eta(x_+^U)}$

We already established the argument with the optimality of μ_0 i.e

$$\mathbb{E}_{v \sim p_v^{\mu_0(x)}}\left[\rho(x,v)\right] \leq \mathbb{E}_{v \sim p_v^{\eta(x)}}\left[\rho(x,v)\right]$$

Stepping one step changes nothing

$$\mathbb{E}_{v \sim p_v^{\mu_0(x_t^v)}}\left[\rho(x_t^v, v_t)\right] \leq \mathbb{E}_{v \sim p_v^{\eta(x_t^v)}}\left[\rho(x_t^v, v_t)\right]$$

So, now

$$V_2(x) \leq \Lambda_2(x) \qquad \forall x$$

Supposing:

$$V_i(x) \leq \Lambda_i(x) \qquad \forall x$$

$$\mu_i(x) = \arg\min_{v} \left\{ \mathbb{E}_{v \sim p_v^{v}}\left[\rho(x,v) + V_i(x_t^v)\right] \right\}$$

$$V_{i+1}(x) := \mathbb{E}_{v \sim p_v^{\mu_i(x)}}\left[\rho(x,v) + V_i(x_t^v)\right]$$

$$\Lambda_{i+1}(x) = \mathbb{E}_{v \sim p_v^{\eta(x)}}\left[\rho(x,v) + \Lambda_i(x_t^v)\right]$$

show: $V_{i+1}(x) \leq \Lambda_{i+1}(x)$

By optimality of μ_i

$$\mathbb{E}_{U\sim P_U^{\mu_i(x)}}\left[g(x,U) + V_i(x_+^U)\right] \leq \mathbb{E}_{U\sim P_U^{\eta(x)}}\left[g(x,U) + V_i(x_+^U\right.$$

$$\mathbb{E}_{U\sim P_U^{\eta(x)}}\left[g(x,U) + V_i(x_x^U)\right] \leq \mathbb{E}_{U\sim P_U^{\eta}}\left[g(x,U) + \Lambda_i(x\right.$$

$V_{i+1}(x)$ $\leq$ $\Lambda_{i+1}(x)$

Proof by induction

Using optimal policy instead of arbitrary

⊙ Showing the boundedness

$$V_i(x) \leq V(x)$$ ← optimum cost-to-go

$$V_i(x) \leq \mathbb{E}_{U_{+K}\sim P_U^{\mu^*(x_{+K})}}\left[\sum_{K=0}^{\infty} g(x_{+K}, U_{+K})\right.$$

$$\mathbb{E}_{U_{+K}\sim P_U^{\mu_i(x_{+K})}}\left[\sum_{K=0}^{i-1} g(x_{+K}, U_{+K}) \,|\, x=x\right]$$

adding non zero rewards

cannot be bigger

⊛ Pointwise increasing

$$V_i(x) \leq V_{i+1}(x)$$

$$V_{i+1}(x) = \mathbb{E}_{U \sim p_U^{\mu_i(x)}}\left[\rho(x, U) + V_i(X_+^U)\right]$$

$$V_i(x) = \mathbb{E}_{U \sim p_U^{\mu_{i-1}(x)}}\left[\rho(x, U) + V_{i-1}(X_+^U)\right]$$

μ_{i-1} is minimizer so

$$\mathbb{E}_{U \sim p_U^{\mu_{i-1}(x)}}\left[\rho(x, U) + V_{i-1}(X_+^U)\right]$$

$$\leq \mathbb{E}_{U \sim p_U^{\mu_i(x)}}\left[\rho(x, U) + V_{i-1}(X_+^U)\right]$$

or any policy for that matter

$$V_i(x) \leq \mathbb{E}_{U \sim p_U^{\mu_i(x)}}\left[\rho(x, U) + V_{i-1}(X_+^U)\right]$$

$$V_i(x) = \mathbb{E}_{U_{t:k} \sim \rho_U^{\mu_{i-1-k}(X_{t:k})}} \left[\sum_{i=0}^{i-1} \rho(X_{t:k}, U_{t:k}) \right] \quad | X = x$$

$$V_{i+1}(x) = \mathbb{E}_{U_{t:k} \sim \rho_U^{\mu_{i-k}(X_{t:k})}} \left[\sum_{i=0}^{i} \rho(X_{t:k}, U_{t:k} \mid X = x \right]$$

Verifying

We have seen this : $i = 2$

$$V_2(x) = \mathbb{E}_{\substack{U \sim \rho_U^{\mu_1(x)} \\ U_+ \sim \rho_U^{\mu_0(X_+^U)}}} \left[\rho(x, U) + \rho(X_+^U, U_+) \right]$$

$\hookrightarrow U_+ \sim \rho_U^{\mu_{i-1-k}(X_{+1})} \Rightarrow 2-1-1 = \rho_U^{\mu_0(X_+)}$

$$V_3(x) = \mathbb{E}_{\substack{U \sim \rho_U^{\mu_2(x)} \\ U_+ \sim \rho_U^{\mu_1(X_+^U)} \\ U_{+2} \sim \rho_U^{\mu_0(X_{+2}^{U_+})}}} \left[\rho(x, U) + \rho(X_+^U, U_+) + \rho(X_{+2}^{U_+}, U_{+2}) \right]$$

$\downarrow$

$$\sum_{k=0}^{i-1} \rho(X_{t:k}, U_{t:k})$$

$\downarrow$

$$U_{t:k} \sim \rho_U^{\mu_{(i-1)-k}}$$

$i = 3$

$V_{i-1}(x_{t+1}) \leq V_{i+1}(x_{t+1})$

$V_{i-1}(x) = V_{i-1}(x_t^u)$

$$V_{i-1}(x_t^u) \leq V_i(x_t^u)$$

$$V_i(x) \leq \mathbb{E}_{u \sim p_u^{\mu_i(x)}} \left[\rho(x,u) + V_{i-1}(x_t^u) \right]$$

$$\leq \mathbb{E}_{u \sim p_u^{\mu_i(x)}} \left[\rho(x,u) + V_i(x_t^u) \right]$$

$$\leq V_{i+1}(x)$$