Two Players Game

The Models

- One player MDP:

$$\forall a \in \mathcal{A}, s \in \mathcal{S} : \pi \xrightarrow{p(s'|s,a)} MC = \left\{ \tau = (s_0, a_0, s_1, \ldots, s_{T-1}, a_{T-1}, s_T) \sim p(s_0) \prod_{t=0}^{T-1} \pi(a_t|s_t) p(s_{t+1}|s_t, a_t) \right\}$$
$$\xrightarrow{r(s,a,s'),\gamma} V^\pi(s_0) = \mathbb{E}\left\{ \sum_{t=0}^{T-1} \gamma^t r(s_t, a_t, s_{t+1}) \Big| s_0 \in \mathcal{S} \right\}$$

- Two players MDP:

$$\forall a, b \in \mathcal{A}, s \in \mathcal{S} : \pi_a, \pi_b \xrightarrow{p(s'|s,a,b)} MC = \Big\{ \tau = (s_0, a_0, b_0, s_1, \ldots, s_{T-1}, a_{T-1}, b_{T-1}, s_T)$$
$$\sim p(s_0) \prod_{t=0}^{T-1} \pi_a(a_t|s_t) \pi_b(b_t|s_t, a_t) p(s_{t+1}|s_t, a_t, b_t) \Big\}$$
$$\xrightarrow{r(s,a,b,s'),\gamma} V^\pi(s_0) = \mathbb{E}\left\{ \sum_{t=0}^{T-1} \gamma^t r(s_t, a_t, b_t, s_{t+1}) \Big| s_0 \in \mathcal{S} \right\}$$

From one-player to two-players

Some Definitions

$$J(\pi_a, \pi_b) = \sum_{s_0} p_0(s_0) \sum_{a_0} \pi_a(a_0|s_0) \sum_{b_0} \pi_b(b_0|s_0, a_0) \sum_{s_1} p(s_1|s_0, a_0, b_0) \Big(r(s_0, a_0, b_0, s_1) +$$
$$\sum_{a_1} \pi_a(a_1|s_1) \sum_{b_1} \pi_b(b_1|s_1, a_1) \sum_{s_2} p(s_2|s_1, a_1, b_1) \gamma \Big(r(s_1, a_1, b_1, s_2) +$$
$$\sum_{a_2} \pi_a(a_2|s_2) \sum_{b_2} \pi_b(b_2|s_2, a_2) \sum_{s_3} p(s_3|s_2, a_2, b_2) \gamma^2 (\cdots) \Big) \Big)$$

$$V(s_0|\pi_a, \pi_b) = \sum_{a_0} \pi_a(a_0|s_0) \sum_{b_0} \pi_b(b_0|s_0, a_0) \sum_{s_1} p(s_1|s_0, a_0, b_0) \Big(r(s_0, a_0, b_0, s_1) +$$
$$\sum_{a_1} \pi_a(a_1|s_1) \sum_{b_1} \pi_b(b_1|s_1, a_1) \sum_{s_2} p(s_2|s_1, a_1, b_1) \gamma \Big(r(s_1, a_1, b_1, s_2) +$$
$$\sum_{a_2} \pi_a(a_2|s_2) \sum_{b_2} \pi_b(b_2|s_2, a_2) \sum_{s_3} p(s_3|s_2, a_2, b_2) \gamma^2 (\cdots) \Big) \Big)$$

$$Q_1(s_0, a_0|\pi_a, \pi_b) = \sum_{b_0} \pi_b(b_0|s_0, a_0) \sum_{s_1} p(s_1|s_0, a_0, b_0) \Big(r(s_0, a_0, b_0, s_1) +$$
$$\sum_{a_1} \pi_a(a_1|s_1) \sum_{b_1} \pi_b(b_1|s_1, a_1) \sum_{s_2} p(s_2|s_1, a_1, b_1) \gamma \Big(r(s_1, a_1, b_1, s_2) +$$
$$\sum_{a_2} \pi_a(a_2|s_2) \sum_{b_2} \pi_b(b_2|s_2, a_2) \sum_{s_3} p(s_3|s_2, a_2, b_2) \gamma^2 (\cdots) \Big) \Big)$$

$$Q_2(s_0, a_0, b_0|\pi_a, \pi_b) = \sum_{s_1} p(s_1|s_0, a_0, b_0) \Bigg(r(s_0, a_0, b_0, s_1) +$$

$$\sum_{a_1} \pi_a(a_1|s_1) \sum_{b_1} \pi_b(b_1|s_1, a_1) \sum_{s_2} p(s_2|s_1, a_1, b_1) \gamma \Bigg(r(s_1, a_1, b_1, s_2) +$$

$$\sum_{a_2} \pi_a(a_2|s_2) \sum_{b_2} \pi_b(b_2|s_2, a_2) \sum_{s_3} p(s_3|s_2, a_2, b_2) \gamma^2 (\cdots) \Bigg) \Bigg)$$

Relationships

$$\begin{cases} J(\pi_a, \pi_b) = \sum_{s \in \mathcal{S}} p_0(s) V(s|\pi_a, \pi_b) \\ V(s|\pi_a, \pi_b) = \sum_a \pi_a(a|s) Q_1(s, a|\pi_a, \pi_b) \\ Q_1(s, a|\pi_a, \pi_b) = \sum_b \pi_b(b|s, a) Q_2(s, a, b|\pi_a, \pi_b) \\ Q_2(s, a, b|\pi_a, \pi_b) = \sum_{s'} p(s'|s, a, b)(r(s, a, b, s') + \gamma V(s')) \end{cases}$$

And $V_{\pi_a,\pi_b} = [V(s|\pi_a, \pi_b)]$, $Q_{1,\pi_a,\pi_b} = [Q_1(s, a|\pi_a, \pi_b)]$, $Q_{2,\pi_a,\pi_b} = [Q_2(s, a, b|\pi_a, \pi_b)]$.

Two Players Bellman Equation

For all $V \in \mathbb{R}^{|\mathcal{S}|}$,

$$T_{\pi_a,\pi_b} V(s) = \sum_a \pi_a(a|s) \sum_b \pi_b(b|s, a) \sum_{s'} p(s'|s, a, b)(r(s, a, b, s') + \gamma V(s')).$$

If T_{π_a,π_b} is contraction mapping then

$$V_{\pi_a,\pi_b} = T_{\pi_a,\pi_b} V_{\pi_a,\pi_b}.$$

In value iteration:

$$J(\pi_a, \pi_b) = \frac{1}{2} \|V_{\pi_a,\pi_b} - T_{\pi_a,\pi_b} V_{\pi_a,\pi_b}\|_\mu^2$$

Matrix Formation

$$V_{\pi_a,\pi_b} = \langle \pi_a, \langle \pi_b, P \rangle \rangle (r + \gamma V_{\pi_a,\pi_b}) \Rightarrow V_{\pi_a,\pi_b} = (I - \gamma \langle \pi_a, \langle \pi_b, P \rangle \rangle)^{-1} \bar{r}$$

$$P = \begin{bmatrix} \alpha_{11}\beta_{111}p_{1111} + \alpha_{11}\beta_{112}p_{1121} + \alpha_{12}\beta_{121}p_{1211} + \alpha_{12}\beta_{122}p_{1221} & \alpha_{11}\beta_{111}p_{1112} + \alpha_{11}\beta_{112}p_{1122} + \alpha_{12}\beta_{121}p_{1212} + \alpha_{12}\beta_{122}p_{1222} \\ \alpha_{21}\beta_{211}p_{2111} + \alpha_{21}\beta_{212}p_{2121} + \alpha_{22}\beta_{221}p_{2211} + \alpha_{22}\beta_{222}p_{2221} & \alpha_{21}\beta_{211}p_{2112} + \alpha_{21}\beta_{212}p_{2122} + \alpha_{22}\beta_{221}p_{2212} + \alpha_{22}\beta_{222}p_{2222} \end{bmatrix}$$

$$\pi_a(\cdot|s_1) = \begin{bmatrix} \pi_a(a_1|s_1) \\ \pi_a(a_2|s_1) \end{bmatrix}$$

$$P(\cdot|\cdot, s_1) = \begin{bmatrix} p(s_1|a_1, s_1) & p(s_1|a_2, s_1) \\ p(s_2|a_1, s_1) & p(s_2|a_2, s_1) \end{bmatrix}$$

$$P(\cdot|s_1) = \pi_a(a_1|s_1) \begin{bmatrix} p(s_1|a_1, s_1) \\ p(s_2|a_1, s_1) \end{bmatrix} + \pi_a(a_2|s_1) \begin{bmatrix} p(s_1|a_2, s_1) \\ p(s_2|a_2, s_1) \end{bmatrix}$$

$$P(\cdot|s_1, a_1) = \pi_b(b_1|s_1, a_1) \begin{bmatrix} p(s_1|s_1, a_1, b_1) \\ p(s_2|s_1, a_1, b_1) \end{bmatrix} + \pi_b(b_2|s_1, a_1) \begin{bmatrix} p(s_1|s_1, a_1, b_2) \\ p(s_2|s_1, a_1, b_2) \end{bmatrix}$$

$$P(\cdot|s_1, a_2) = \pi_b(b_1|s_1, a_2) \begin{bmatrix} p(s_1|s_1, a_2, b_1) \\ p(s_2|s_1, a_2, b_1) \end{bmatrix} + \pi_b(b_2|s_1, a_2) \begin{bmatrix} p(s_1|s_1, a_2, b_2) \\ p(s_2|s_1, a_2, b_2) \end{bmatrix}$$

Two Players Optimal Bellman Equation

For all $V \in \mathbb{R}^{|\mathcal{S}|}$,

$$T_{\pi_b} V(s) = \max_{\pi_a(\cdot|s)} \sum_a \pi_a(a|s) \sum_b \pi_b(b|s,a) \sum_{s'} p(s'|s,a,b)(r(s,a,b,s') + \gamma V(s')).$$

In original optimal Bellman equation,

$$V_{*,\pi_b} = T_{\pi_b} V_{*,\pi_b}.$$

We denote $\pi_a^*(\pi_b)$ that satisfies $V_{\pi_a^*(\pi_b),\pi_b} = V_{*,\pi_b}$.