Last Time

Simple Games, Nash Equilibrium

attacker \ defender	climb	descend
climb	(3, -5)	(-1, 0)
descend	(0, -1)	(4, -4)

	π(climb)	π(descend)	U
attacker	0.375,	0.625	1.5
defender	0.625,	0.375	-2.5

This Time

Sequential Games

- Zero Sum — $R^1(\vec{a}) = -R^2(\vec{a})$
- Cooperative Game — $R^i(\vec{a}) = R(\vec{a})$
- General Sum

Markov Game

(I, S, A, T, R, γ)

A: joint action

$T(s' \mid s, \vec{a})$

Stochastic / mixed policies might be needed to describe equilibria

$\pi^i(a \mid s)$ $\quad \vec{\pi}(s)$

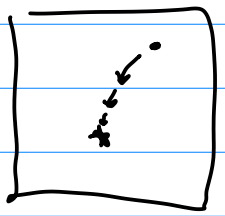
Best Response

Fix policies of other agents ($\vec{\pi}^{-i}(s)$ is fixed)

↳ Results in MDP

Nash Equilibrium

Will Value Iteration Converge?



$$\underset{\pi, U}{\text{minimize}} \quad \sum_{i \in I} \sum_{s} \left(U^i(s) - Q^i(s, \pi(s)) \right)$$

subject to

$$U^i(s) \geq Q^i(s, a^i, \vec{\pi}^{-i}(s)) \quad \forall\, i, s, a^i$$

$$\sum_{a^i} \pi^i(a^i \mid s) = 1 \quad \forall\, i, s$$

$$\pi^i(a^i \mid s) \geq 0 \quad \forall\, i, s, a^i$$

where $Q^i(s, \vec{\pi}(s)) = R^i(s, \vec{\pi}(s)) + \gamma \sum_{s'} T(s' \mid s, \vec{\pi}(s)) U^i(s')$

Fictitious Play

$\pi^i(a^i|s) \propto N^i(j, a^i, s)$

Compute best response (solve MDP) periodically

Gadient Ascent

$$\frac{\partial U^{\pi_{t,i}}(s)}{\partial \pi_t^i(a^i|s)} = Q^{\pi_{t,i}}(s_t, a_t^i, a_t^{-i})$$

Nash Q-learning

$\bar{Q}(s, \bar{a})$ joint action-value function

At every step

Compute Nash $\bar{a}'$

$$\bar{U}(s') = \sum_{\bar{a}'} \bar{Q}(s', \bar{a}') \prod_{j \in I} \pi^j(a^j)$$

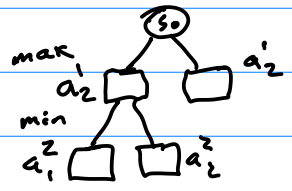
s $\bar{Q}(s, \bar{a})$

	a_1^2	a_2^2
a_1^1	Q^1, Q^2	Q^1, Q^2
a_2^1	Q^1, Q^2	Q^1, Q^2

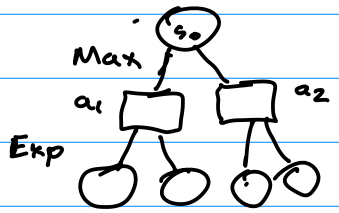
$$\bar{Q}(s, \bar{a}) \leftarrow \bar{Q}(s, \bar{a}) + \alpha\left(\bar{R}(s, \bar{a}) + \gamma\, \bar{U}(s') - \bar{Q}(s, \bar{a})\right)$$

Zero-Sum

Zero Sum Deterministic



MDP



State Uncertainty

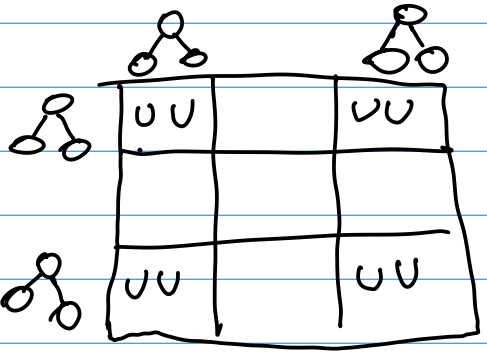
POMG

"POSG"

"Extensive-Form Game"

$(I, S, A, O, T, Z, R, \gamma)$

Belief updates "not possible"



Dynamic Programing

similar to POMDP value iteration

Differential Game

$$\dot{x} = f(t, x, u_1, \dots, u_N)$$

$$J_i = \int_0^T g_i(t, x, u_1, \dots, u_N)\, dt$$

Hamilton
Jacobi
Reachability

$d \in D$

Markov Game

$$x_{k+1} = G(x_k, \bar{a}, w)$$