

Deep Reinforcement Learning

Agents learn Predatory Pricing

Motivation

- Complex economic models have been infesible to analyse until recently
 - **Reinforcement Learning** (RL) can solve complex decision making problems
- ⇒ What if RL controls economic agents?

Dynamic Oligopoly Model

Agents are companies competing in an oligopoly over multiple rounds:

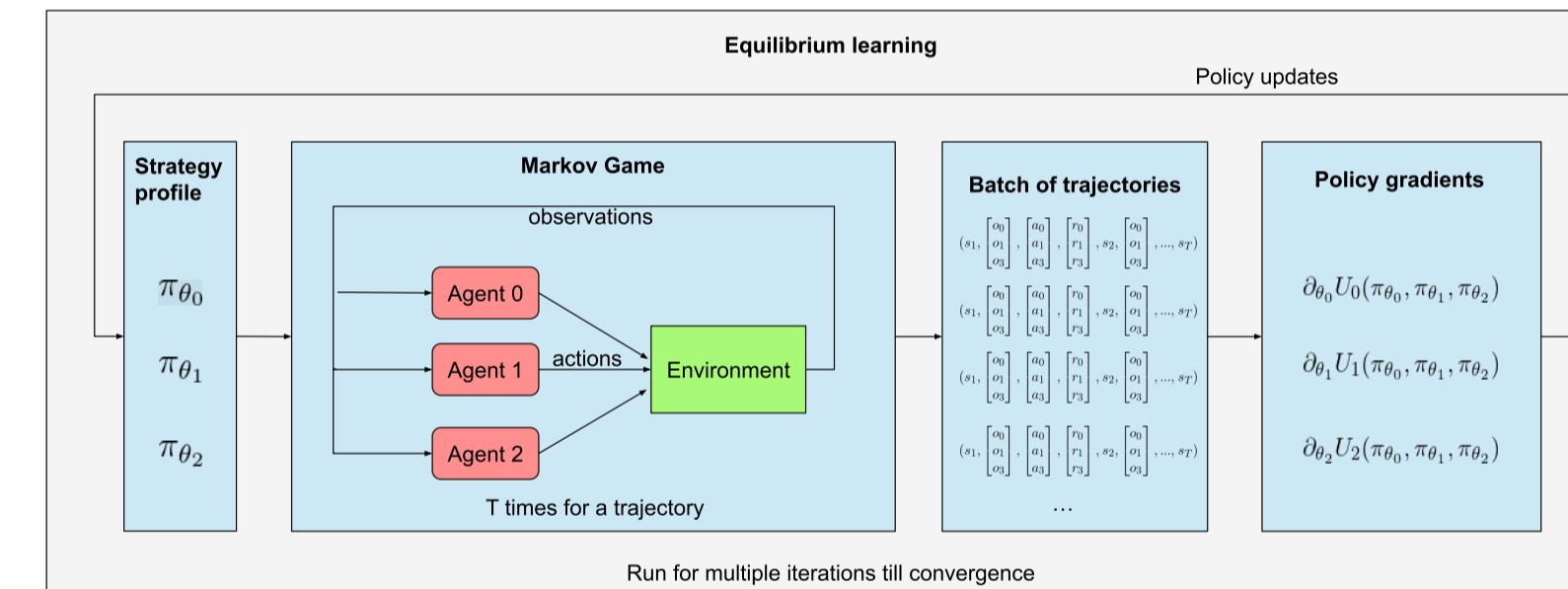
Require:
 Set of agents $\mathcal{N} = \{1, \dots, N\}$
 Number of rounds T
 For each agent i : initial demand D_1^i , unit production cost c_i , policy π_i , observation function Φ_i
for $t = 1, 2, \dots, T$ **do**
 for $i \in \mathcal{N}$ **do**
 i observes $o_t^i = \Phi_i(s_t) = \Phi_i(t, D_t^1, \dots, D_t^N)$
 i selects a price $p_t^i \sim \pi_i(o_t^i)$
 i sells quantity $D_t^i - p_t^i$
 i receives reward $r_t^i = (p_t^i - c_i)(D_t^i - p_t^i)$
 end for
 Compute the average price as $\bar{p}_t = \frac{1}{N} \sum_{j \in \mathcal{N}} p_t^j$
 for $i \in \mathcal{N}$ **do**
 Compute the price difference $\Delta p_t^i = p_t^i - \bar{p}_t$
 Transition demand to $D_{t+1}^i = D_t^i - \Delta p_t^i$
 Optionally, drop out i if $D_{t+1}^i < c_i$ (see Eq. (2))
 end for
end for
 Reward each agent i with $U_i = \sum_{t=1}^T r_t^i$

Equilibrium Learning

We search **Nash equilibria** as fixed points of the game where no player gains by deviating:

$$\sup_{\pi_i \in \Sigma_i} U_i(\pi_i, \pi_{-i}^*) - U_i(\pi_i^*, \pi_{-i}) \leq \varepsilon \quad \forall i \in \mathcal{N}$$

Each company is controlled by an RL agent aiming to maximize its profit:



We use REINFORCE (or its variants) to update all agents' policies simultaneously:

$$\theta_i \leftarrow \theta_i + \alpha \nabla_{\theta_i} U_i(\pi_{\theta_i}, \{\pi_{\theta_j}\}_{j \in \mathcal{N} \setminus \{i\}})$$

We hope this procedure converges to a Nash equilibrium, as measured by:

$$\mathcal{L}_{\text{bf}} = \sum_{i \in \mathcal{N}} \sup_{\pi_i \in \Sigma_i^K} U_i(\pi_i, \pi_{-i}) - U_i(\pi_i, \pi_{-i})$$

