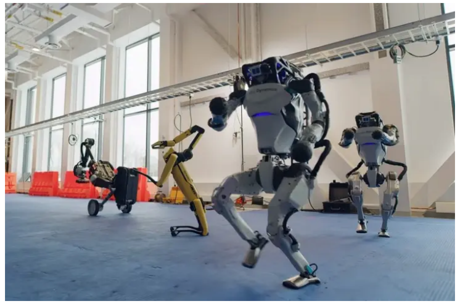
Markov games 🤖🤖🎲🎲

Multiple agents interact with each other in a dynamically changing environment.

Motivation

Auctions

Self-driving cars



Robotics

E-Sports

Mathematical definition 🤔💭🔢

A Markov game is a tuple $\Gamma = (\mathcal{S}, \mathcal{N}, \mathcal{A}, \{r_i\}, P, \gamma, \mu)$:

- a finite number of states $\mathcal{S}$

Mathematical definition 🤔💭🔢

A Markov game is a tuple $\Gamma = (\mathcal{S}, \mathcal{N}, \mathcal{A}, \{r_i\}, P, \gamma, \mu)$:

- a finite number of states $\mathcal{S}$
- a finite number of players $\mathcal{N}$

Mathematical definition 🤔💭🔢

A Markov game is a tuple $\Gamma = (\mathcal{S}, \mathcal{N}, \mathcal{A}, \{r_i\}, P, \gamma, \mu)$:

- a finite number of states $\mathcal{S}$
- a finite number of players $\mathcal{N}$
- each agent $i \in \mathcal{N}$ gets a finite number of actions $\mathcal{A}_i$

Mathematical definition 🤔💭🔢

A Markov game is a tuple $\Gamma = (\mathcal{S}, \mathcal{N}, \mathcal{A}, \{r_i\}, P, \gamma, \mu)$:

- a finite number of states $\mathcal{S}$
- a finite number of players $\mathcal{N}$
- each agent $i \in \mathcal{N}$ gets a finite number of actions $\mathcal{A}_i$
- a reward function $r_i : \mathcal{S} \times \mathcal{A}_1 \times \cdots \mathcal{A}_n \to [-1, 1]$

Mathematical definition 🤔💭🔢

A Markov game is a tuple $\Gamma = (\mathcal{S}, \mathcal{N}, \mathcal{A}, \{r_i\}, P, \gamma, \mu)$:

- a finite number of states $\mathcal{S}$
- a finite number of players $\mathcal{N}$
- each agent $i \in \mathcal{N}$ gets a finite number of actions $\mathcal{A}_i$
- a reward function $r_i : \mathcal{S} \times \mathcal{A}_1 \times \cdots \mathcal{A}_n \to [-1, 1]$
- a probability transition function $P : \mathcal{S} \times \mathcal{A}_1 \times \cdots \mathcal{A}_n \to \Delta(\mathcal{S})$

Mathematical definition 🤔💭🔢

A Markov game is a tuple $\Gamma = (\mathcal{S}, \mathcal{N}, \mathcal{A}, \{r_i\}, P, \gamma, \mu)$:

- a finite number of states $\mathcal{S}$
- a finite number of players $\mathcal{N}$
- each agent $i \in \mathcal{N}$ gets a finite number of actions $\mathcal{A}_i$
- a reward function $r_i : \mathcal{S} \times \mathcal{A}_1 \times \cdots \mathcal{A}_n \to [-1, 1]$
- a probability transition function $P : \mathcal{S} \times \mathcal{A}_1 \times \cdots \mathcal{A}_n \to \Delta(\mathcal{S})$
- a discount factor $\gamma \in [0, 1)$

Mathematical definition 🤔💭🔢

A Markov game is a tuple $\Gamma = (\mathcal{S}, \mathcal{N}, \mathcal{A}, \{r_i\}, P, \gamma, \mu)$:

- a finite number of states $\mathcal{S}$
- a finite number of players $\mathcal{N}$
- each agent $i \in \mathcal{N}$ gets a finite number of actions $\mathcal{A}_i$
- a reward function $r_i : \mathcal{S} \times \mathcal{A}_1 \times \cdots \mathcal{A}_n \to [-1, 1]$
- a probability transition function $P : \mathcal{S} \times \mathcal{A}_1 \times \cdots \mathcal{A}_n \to \Delta(\mathcal{S})$
- a discount factor $\gamma \in [0, 1)$
- $\mu \in \Delta(\mathcal{S})$ an initial state distribution.

Policy and value function 🧠📈

The objective of each agent i is to maximize their own **value function**:

$$V_i^\pi(\mu) = \mathbb{E}_\pi[r_i^{(1)} + \gamma r_i^{(2)} + \gamma^2 r_i^{(3)} + \dots]$$
$$= \mathbb{E}_\pi\Big[\sum_{t=0}^{\infty} \gamma^t r_i(s^{(t)}, a_1^{(t)}, \dots, a_n^{(t)}) \mid s_0 \sim \mu\Big].$$

Where each agent i controls their own **policy**, *i.e.*,

$$\pi_i : \mathcal{S} \to \Delta(\mathcal{A}_i).$$

Also, the policy profile is denoted $\pi = (\pi_1, \dots, \pi_n)$.

Existence of Nash equilibria in n-player Markov games ⚖️

> **Theorem.** (Fink 1964) There always exists a Nash equilibrium for every Markov game Γ.

Existence of Nash equilibria in n-player Markov games ⚖️

> **Theorem.** (Fink 1964) There always exists a Nash equilibrium for every Markov game Γ.

Equivalently, there exists $\pi^* = (\pi_1^*, \ldots, \pi_n^*)$:

$$V_i^{\pi^*}(\mu) \geq V_i^{\pi_i', \pi_{-i}^*}(\mu), \ \forall \pi_i'.$$

Markov games are at least as hard as normal-form games

- Let the time horizon be equal to 1 and only one possible state in the game.
- Then, the Markov game becomes a normal-form game.
- Hence, they cannot be *easier* than normal-form games.

Some tractable instances of Markov games 💻✖️🧮⏳

- Two-player zero-sum games

Some tractable instances of Markov games 💻❌🧮⏳

- Two-player zero-sum games
- Markov potential games

Two-player zero-sum Markov games 🐱 🐭

- a Markov game $\Gamma(\mathcal{N}, \mathcal{A}, \{r_i\}_{i \in \mathcal{N}}, P, \gamma, \mu)$,

Two-player zero-sum Markov games 🐱 🐭

- a Markov game $\Gamma(\mathcal{N}, \mathcal{A}, \{r_i\}_{i\in\mathcal{N}}, P, \gamma, \mu)$,
- two players $\mathcal{N} = \{1, 2\}$,

Two-player zero-sum Markov games 🐱 🐭

- a Markov game $\Gamma(\mathcal{N}, \mathcal{A}, \{r_i\}_{i\in\mathcal{N}}, P, \gamma, \mu)$,
- two players $\mathcal{N} = \{1, 2\}$,
- two finite action set $\mathcal{A}, \mathcal{B}$,

Two-player zero-sum Markov games 🐱 🐭

- a Markov game $\Gamma(\mathcal{N}, \mathcal{A}, \{r_i\}_{i\in\mathcal{N}}, P, \gamma, \mu)$,
- two players $\mathcal{N} = \{1, 2\}$,
- two finite action set $\mathcal{A}, \mathcal{B}$,
- the sum of the rewards is always equal to 0,
  i.e., $r(s, a, b) = r_2(s, a, b) = -r_1(s, a, b)$.

Two-player zero-sum Markov games 🐱 🐭

- a Markov game $\Gamma(\mathcal{N}, \mathcal{A}, \{r_i\}_{i\in\mathcal{N}}, P, \gamma, \mu)$,
- two players $\mathcal{N} = \{1, 2\}$,
- two finite action set $\mathcal{A}, \mathcal{B}$,
- the sum of the rewards is always equal to 0,
  i.e., $r(s, a, b) = r_2(s, a, b) = -r_1(s, a, b)$.

Conventions

- We call player 2 the maximizer and player 1 the minimizer.
- Define the value function of the maximizer $V^{\pi_1, \pi_2}(s)$.

A crucial property

> **Theorem.** (Shapley 1953): In any two-player zero-sum game:

$$\min_{\pi_1} \max_{\pi_2} V^{\pi_1,\pi_2}(\mu) = \max_{\pi_2} \min_{\pi_1} V^{\pi_1,\pi_2}(\mu).$$

A crucial property

> **Theorem.** (Shapley 1953): In any two-player zero-sum game:

$$V^* = \min_{\pi_1} \max_{\pi_2} V^{\pi_1,\pi_2}(\mu) = \max_{\pi_2} \min_{\pi_1} V^{\pi_1,\pi_2}(\mu).$$

- The "duality gap" is equal to zero. (Remember two-pl. normal-form games!)

A crucial property

> **Theorem.** (Shapley 1953): In any two-player zero-sum game:

$$V^* = \min_{\pi_1} \max_{\pi_2} V^{\pi_1, \pi_2}(\mu) = \max_{\pi_2} \min_{\pi_1} V^{\pi_1, \pi_2}(\mu).$$

- The "duality gap" is equal to zero. (Remember two-pl. normal-form games!)
- It does not matter who commits first to a policy.

Proof.

- Define the operator on matrices $\mathrm{val}(\cdot)$:
  - given a matrix, it outputs the minimax value of that matrix.
  - *e.g.* $\mathrm{val}\left(\begin{bmatrix}-1, 1\\ 1, -1\end{bmatrix}\right) = 0.$

Proof. (cont.)

- Initialize a vector $v^{(0)} \in \mathbb{R}^{|\mathcal{S}|}$ arbitrarily.
- We define the following iterative process:

$$v^{(k+1)}(s) = \mathrm{val}\left(r(s,\cdot,\cdot) + \gamma \sum_{s'} P(s'|s,\cdot,\cdot) v^{(k)}(s') \right), \forall s \in \mathcal{S}.$$

- For shorthand, we define the operator $\mathcal{T}$:

$$v^{(k+1)} = \mathcal{T} v^{(k)}.$$

Proof. (cont.)

- Initialize a vector $v^{(0)} \in \mathbb{R}^{|\mathcal{S}|}$ arbitrarily.
- We define the following iterative process:

$$v^{(k+1)}(s) = \mathrm{val}\left(r(s,\cdot,\cdot) + \gamma \sum_{s'} P(s'|s,\cdot,\cdot)v^{(k)}(s')\right), \forall s \in \mathcal{S}.$$

- For shorthand, we define the operator $\mathcal{T}$:

$$v^{(k+1)} = \mathcal{T}v^{(k)}.$$

Proof. (cont.)

- Let $w = \mathcal{T}v$.
- Observe that:

$$\|\mathcal{T}w - \mathcal{T}v\|_\infty \leq \max_s \left| \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot) w(s') \right) - \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot) v(s') \right) \right|$$

Proof. (cont.)

- Let $w = \mathcal{T}v$.
- Observe that:

$$\begin{aligned}\|\mathcal{T}w - \mathcal{T}v\|_\infty &\leq \max_s \left|\mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot)w(s')\right) - \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot)v(s')\right)\right| \\ &\leq \max_s \max_{a,b} \left|\gamma \sum P(s'|s,a,b)w(s') - \gamma \sum P(s'|s,a,b)v(s')\right|\end{aligned}$$

Proof. (cont.)

- Let $w = \mathcal{T}v$.
- Observe that:

$$
\begin{aligned}
\|\mathcal{T}w - \mathcal{T}v\|_\infty &\le \max_s \left| \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot)w(s')\right) - \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot)v(s')\right) \right| \\
&\le \max_s \max_{a,b} \left| \gamma \sum P(s'|s,a,b)w(s') - \gamma \sum P(s'|s,a,b)v(s') \right| \\
&\le \gamma \max_{s,a,b} \left| P(\cdot|s,a,b) \right| \max_{s'} \left| w(s') - v(s') \right|
\end{aligned}
$$

Proof. (cont.)

- Let $w = \mathcal{T}v$.
- Observe that:

$$
\begin{aligned}
\|\mathcal{T}w - \mathcal{T}v\|_\infty &\le \max_s \left| \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot) w(s') \right) - \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot) v(s') \right) \right| \\
&\le \max_s \max_{a,b} \left| \gamma \sum P(s'|s,a,b) w(s') - \gamma \sum P(s'|s,a,b) v(s') \right| \\
&\le \gamma \max_{s,a,b} \left| P(\cdot|s,a,b) \right| \max_{s'} \left| w(s') - v(s') \right| \\
&\le \gamma \|w - v\|_\infty
\end{aligned}
$$

Proof. (cont.)

- Let $w = \mathcal{T}v$.
- Observe that:

$$
\begin{aligned}
\|\mathcal{T}w - \mathcal{T}v\|_\infty &\le \max_s \left| \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot)w(s')\right) - \mathrm{val}\left(\gamma \sum P(s'|s,\cdot,\cdot)v(s')\right) \right| \\
&\le \max_s \max_{a,b} \left| \gamma \sum P(s'|s,a,b)w(s') - \gamma \sum P(s'|s,a,b)v(s') \right| \\
&\le \gamma \max_{s,a,b} \left| P(\cdot|s,a,b) \right| \max_{s'} \left| w(s') - v(s') \right| \\
&\le \gamma \|w - v\|_\infty = \gamma \|\mathcal{T}v - v\|_\infty .
\end{aligned}
$$

Proof. (cont.)

- Hence,

$$\|\mathcal{T}^2 v - \mathcal{T} v\|_\infty \leq \gamma \|\mathcal{T} v - v\|, \text{ for all } v \in \mathbb{R}^{|\mathcal{S}|}.$$

- *I.e.*, the operator $\mathcal{T}$ is a contraction
- From Banach's fixed point theorem, $\mathcal{T}$ has a unique fixed point!
- This unique fixed point, $\mathcal{T} V^* = V^*$,

$$V^* = \min_{\pi_1} \max_{\pi_2} V^{\pi_1,\pi_2}(\mu) = \max_{\pi_2} \min_{\pi_1} V^{\pi_1,\pi_2}(\mu).$$