Paths to Equilibrium in Games

B. Yongacoglu[1] G. Arslan[2] L. Pavel[1] S. Yuksel[3]

[1]University of Toronto

[2]University of Hawaii at Manoa

[3]Queen's University

Multi-Agent Reinforcement Learning

Multi-agent learning in games:

- Shared environment
- Coupled rewards
- Iterative strategy revision
  - experiment $\rightarrow$ assess $\rightarrow$ revise $\rightarrow$ experiment $\rightarrow$ ...

$\Rightarrow$ Non-stationary learning problem with challenging analysis

Goal

Identify structure that can help design and analyze algorithms for games.

UNIVERSITY OF TORONTO

Win-Stay, Lose-Shift Algorithms

Prior work on learning in games:

- Deep analysis of particular algorithms.
- Structural (im)possibility results for dynamics in the strategy space.

Our motivation: understand *win-stay*, *lose-shift algorithms*.

- Generalize algorithms driven by best responding
- Incorporate random search ⇒ irregular strategy dynamics

Q: What are the limitations of such algorithms?

Model: Finite Normal-Form Games

A game $\Gamma = \left(n, \mathbf{X}, \{R^i\}_{i=1}^n\right)$ is played as follows:

- Player i selects a strategy $x^i \in \mathcal{X}^i$, for $i = 1, 2, \ldots, n$
- The strategy profile is denoted $\mathbf{x} = (x^i)_{i=1}^n$.
- Player i receives reward $R^i(\mathbf{x}) = R^i(x^i, \mathbf{x}^{-i})$.
- $x^i_\star \in \mathcal{X}^i$ is a best response to $\mathbf{x}^{-i}$ if it maximizes $R^i(\cdot, \mathbf{x}^{-i})$ over $\mathcal{X}^i$.
- $\mathrm{BR}^i(\mathbf{x}^{-i})$ denotes player i's set of best responses to $\mathbf{x}^{-i}$.

If $x^i \in \mathrm{BR}^i(\mathbf{x}^{-i})$, we say that player i is "satisfied" at $(x^i, \mathbf{x}^{-i})$.
If $x^i \notin \mathrm{BR}^i(\mathbf{x}^{-i})$, we say that player i is "unsatisfied" at $(x^i, \mathbf{x}^{-i})$.

Win-Stay, Lose-Shift Algorithms (continued)

Win-stay, *lose-shift algorithms* generalize best-response updating:

Best-response updating:

$$x^i_{t+1} = \begin{cases} x^i_t, & \text{if } x^i_t \in \mathcal{B}^i_t \\ \text{some } x^i_\star \in \mathcal{B}^i_t, & \text{else.} \end{cases}$$

where $\mathcal{B}^i_t = \mathrm{BR}^i(\mathbf{x}^{-i}_t)$

Win-Stay, Lose-Shift Algorithms (continued)

Win-stay, *lose-shift algorithms* generalize best-response updating:

Best-response updating:

$$x^i_{t+1} = \begin{cases} x^i_t, & \text{if } x^i_t \in \mathcal{B}^i_t \\ \text{some } x^i_\star \in \mathcal{B}^i_t, & \text{else.} \end{cases}$$

where $\mathcal{B}^i_t = \mathrm{BR}^i(\mathbf{x}^{-i}_t)$

Win-stay, lose-shift updating:

$$x^i_{t+1} = \begin{cases} x^i_t, & \text{if } x^i_t \in \mathcal{B}^i_t \\ ?, & \text{else.} \end{cases}$$

where '?' is a design choice

UNIVERSITY OF TORONTO

Win-Stay, Lose-Shift Algorithms (continued)

Win-stay, lose-shift algorithms generalize best-response updating:

Best-response updating:

$$x^i_{t+1} = \begin{cases} x^i_t, & \text{if } x^i_t \in \mathcal{B}^i_t \\ \text{some } x^i_\star \in \mathcal{B}^i_t, & \text{else.} \end{cases}$$

where $\mathcal{B}^i_t = \mathrm{BR}^i(\mathbf{x}^{-i}_t)$

Win-stay, lose-shift updating:

$$x^i_{t+1} = \begin{cases} x^i_t, & \text{if } x^i_t \in \mathcal{B}^i_t \\ \textbf{?}, & \text{else.} \end{cases}$$

where '?' is a design choice

Advantages of Win-Stay, Lose-Shift Algorithms:

- Exploration: **?** may be random experimentation.
- Fixed points: equilibria (and only equilibria) are invariant.
- Breaking cycles: rigidly requiring $x^i_{t+1} \in \mathcal{B}^i_t$ can cause cycles.

UNIVERSITY OF TORONTO

Satisficing Paths

Definition: Satisficing Paths

A sequence of strategy profiles $\{\mathbf{x}_t\}_{t\geq 1}$ is called a *satisficing path* if

$$x_t^i \in \mathrm{BR}^i(\mathbf{x}_t^{-i}) \implies x_{t+1}^i = x_t^i \qquad \forall i \in [n], t \geq 1.$$

Note: any *Win-Stay, Lose-Shift* algorithm will give rise to a satisficing path.

Satisficing Paths

Definition: Satisficing Paths

A sequence of strategy profiles $\{\mathbf{x}_t\}_{t\geq 1}$ is called a *satisficing path* if

$$x_t^i \in \mathrm{BR}^i(\mathbf{x}_t^{-i}) \implies x_{t+1}^i = x_t^i \qquad \forall i \in [n], t \geq 1.$$

Note: any *Win-Stay, Lose-Shift* algorithm will give rise to a satisficing path.

Question: for a game Γ and starting strategy profile $\mathbf{x}_1$, can we guarantee that a satisficing path from $\mathbf{x}_1$ to some Nash equilibrium of Γ always exists?

Alternatively: can play be driven to equilibrium by switching only the strategies of agents that are unsatisfied?

UNIVERSITY OF TORONTO

Examples of Satisficing Paths in *Rock Paper Scissors*

Legend

$$\mathbf{x} = \left(\begin{bmatrix} \theta_r^1 \\ \theta_p^1 \\ \theta_s^1 \end{bmatrix}, \begin{bmatrix} \theta_r^2 \\ \theta_p^2 \\ \theta_s^2 \end{bmatrix} \right),$$

θ_a^i = prob. player i plays a,

$a \in \{\mathrm{Rock}, \mathrm{Paper}, \mathrm{Scissors}\}.$

$$\mathbf{x} = \left(\begin{bmatrix} \theta_r^i \\ \theta_p^i \\ \theta_s^i \end{bmatrix}, \begin{bmatrix} \theta_r^j \\ \theta_p^j \\ \theta_s^j \end{bmatrix} \right),$$

x^i green: i satisfied at $\mathbf{x}$.

x^i orange: j unsatisfied at $\mathbf{x}$.

Examples of Satisficing Paths in *Rock Paper Scissors*

Ex. 1: Random experimentation when unsatisfied

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \to \left(\begin{bmatrix} 1/2 \\ 1/3 \\ 1/6 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \to \left(\begin{bmatrix} 0.92 \\ 0.01 \\ 0.07 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right) \to \cdots$$

Legend

$$\mathbf{x} = \left(\begin{bmatrix} \theta^1_r \\ \theta^1_p \\ \theta^1_s \end{bmatrix}, \begin{bmatrix} \theta^2_r \\ \theta^2_p \\ \theta^2_s \end{bmatrix} \right),$$

θ^i_a = prob. player i plays a,

$a \in \{\mathrm{Rock}, \mathrm{Paper}, \mathrm{Scissors}\}$.

$$\mathbf{x} = \left(\begin{bmatrix} \theta^i_r \\ \theta^i_p \\ \theta^i_s \end{bmatrix}, \begin{bmatrix} \theta^j_r \\ \theta^j_p \\ \theta^j_s \end{bmatrix} \right),$$

x^i green: i satisfied at $\mathbf{x}$.

x^i orange: j unsatisfied at $\mathbf{x}$.

UNIVERSITY OF TORONTO

Examples of Satisficing Paths in *Rock Paper Scissors*

Ex. 1: Random experimentation when unsatisfied

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right) \to \left(\begin{bmatrix}1/2\\1/3\\1/6\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right) \to \left(\begin{bmatrix}0.92\\0.01\\0.07\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right) \to \cdots$$

Ex. 2: Best-responding (cycles)

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right) \to \left(\begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right) \to \left(\begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}\right) \to \cdots$$

Legend

$$\mathbf{x} = \left(\begin{bmatrix}\theta^1_r\\\theta^1_p\\\theta^1_s\end{bmatrix}, \begin{bmatrix}\theta^2_r\\\theta^2_p\\\theta^2_s\end{bmatrix}\right),$$

θ^i_a = prob. player i plays a,

$a \in \{\mathrm{Rock}, \mathrm{Paper}, \mathrm{Scissors}\}$.

$$\mathbf{x} = \left(\begin{bmatrix}\theta^i_r\\\theta^i_p\\\theta^i_s\end{bmatrix}, \begin{bmatrix}\theta^j_r\\\theta^j_p\\\theta^j_s\end{bmatrix}\right),$$

x^i green: i satisfied at $\mathbf{x}$.

x^i orange: j unsatisfied at $\mathbf{x}$.

UNIVERSITY OF TORONTO

Examples of Satisficing Paths in *Rock Paper Scissors*

Ex. 1: Random experimentation when unsatisfied

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right)\to\left(\begin{bmatrix}1/2\\1/3\\1/6\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right)\to\left(\begin{bmatrix}0.92\\0.01\\0.07\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right)\to\cdots$$

Ex. 2: Best-responding (cycles)

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right)\to\left(\begin{bmatrix}0\\0\\1\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix}\right)\to\left(\begin{bmatrix}0\\0\\1\end{bmatrix},\begin{bmatrix}1\\0\\0\end{bmatrix}\right)\to\cdots$$

Ex. 3: Updates that increase the number of unsatisfied players + seek Nash equilibrium when all players are unsatisfied

$$\left(\begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix},\begin{bmatrix}0\\1/2\\1/2\end{bmatrix}\right)\to\left(\begin{bmatrix}0\\1/2\\1/2\end{bmatrix},\begin{bmatrix}0\\1/2\\1/2\end{bmatrix}\right)\to\left(\begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix},\begin{bmatrix}1/3\\1/3\\1/3\end{bmatrix}\right)$$

Legend

$$\mathbf{x}=\left(\begin{bmatrix}\theta^1_r\\\theta^1_p\\\theta^1_s\end{bmatrix},\begin{bmatrix}\theta^2_r\\\theta^2_p\\\theta^2_s\end{bmatrix}\right),$$

θ^i_a = prob. player i plays a,

$a\in\{\mathrm{Rock},\mathrm{Paper},\mathrm{Scissors}\}$.

$$\mathbf{x}=\left(\begin{bmatrix}\theta^i_r\\\theta^i_p\\\theta^i_s\end{bmatrix},\begin{bmatrix}\theta^j_r\\\theta^j_p\\\theta^j_s\end{bmatrix}\right),$$

x^i green: i satisfied at $\mathbf{x}$.

x^i orange: j unsatisfied at $\mathbf{x}$.

UNIVERSITY OF TORONTO

Main Result on Path Connectivity

Theorem 1

Any finite normal-form game Γ has the satisficing paths property.

(That is, from any initial strategy profile $\mathbf{x}_1$, there exists a satisficing path connecting $\mathbf{x}_1$ to a Nash equilibrium of Γ.)

Insights to leverage:

- Satisfied players are constrained, but unsatisfied players are free
- *Trying to increase the number of satisfied players* (by switching unsatisfied player strategies to best responses) may cause cycling
- When all players are unsatisfied, the satisficing path may proceed to any successor strategy – including jumping to equilibrium in one step.

UNIVERSITY OF TORONTO

Proof Sketch

Beginning at arbitrary $\mathbf{x}_1$, we analytically construct a path to some equilibrium.

Strategy:

- At each iteration t, select $\mathbf{x}_{t+1}$ so the set of unsatisfied players grows.
- When the set of unsatisfied players is **maximal**, this process ends with $\mathbf{x}_k$.
  - If player i is unsatisfied at $\mathbf{x}_k$, free to switch.
  - If player j is satisfied at $\mathbf{x}_k$, must use $x^j_{k+1} = x^j_k$.
- Find an equilibrium for a related subgame (involves only unsatisfied players).
  $\rightarrow$ Choose $\mathbf{x}_{k+1}$ to switch strategies of unsatisfied players to this.
- **(Key) Lemma:** $\mathbf{x}_{k+1}$ is a Nash equilibrium of Γ.
  - Players satisfied at $\mathbf{x}_k$ could (in principle) be unsatisfied at $\mathbf{x}_{k+1}$.
  - Requires analysis of indifference conditions for players satisfied at $\mathbf{x}_k$.

UNIVERSITY OF TORONTO

Conclusion

Summary

- We studied satisficing paths, with the aim of better understanding win-stay, lose-shift algorithms for multi-agent reinforcement learning.
- We showed that satisficing paths to equilibrium always exist in finite normal-form games.

Related open questions

- ϵ-satisficing, defined by ϵ-best-response constraint
- Extension to constrained subsets of strategies
- Extension to Markov games