Convergence to Pure-Strategy Nash Equilibria Under Simple Learning Rules and Selection of Resilient Pure-Strategy Nash Equilibria

Richard J. La
joint work with Siddharth Pal

Department of ECE & ISR

University of Maryland, College Park

hyongla@umd.edu

April 2, 2015

Overview

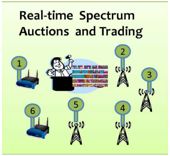
- Background
- Setup
- Classification of games
- Overview of existing literature
- Main results
  - Algorithm #1 - Generalized Better Reply Path Algorithm
  - Algorithm #2 - Simple Experimentation with Monitoring
- Future directions

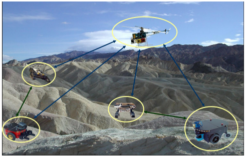
Overview

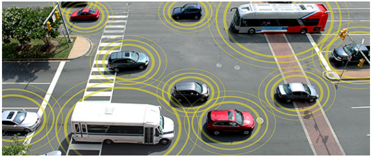
- **Background**
- Setup
- Classification of games
- Overview of existing literature
- Main results
  - Algorithm #1 - Generalized Better Reply Path Algorithm
  - Algorithm #2 - Simple Experimentation with Monitoring
- Future directions

Background (1)

- Increasing interest in application of game-theoretic framework to engineering systems
  - Communication networks
  - Distributed control and systems
  - Transportation networks and systems
  - Supply chain and inventory management







Background (2)

- Game theory is **NOT** about ...

- **Question:** Why is that only the men look angry and not enjoying the game?

Background (3)

- **Game theory** – Study of rational decision making and/or strategic interactions among multiple rational decision makers (“**players**”’) in situations of conflict and/or cooperation
    - Decision – choice of which action/strategy to take based on available information
    - Consequences of decisions captured by **payoffs** or **utilities**
    - Implicit assumption **–** **interdependency** in payoffs/utilities among players through choices
- **Game** – a *mathematical model* that approximates complicated reality
    - Many different types of games
    - Suitable game depends on many factors
    - **Leaves out many details of the reality**

Background (4)

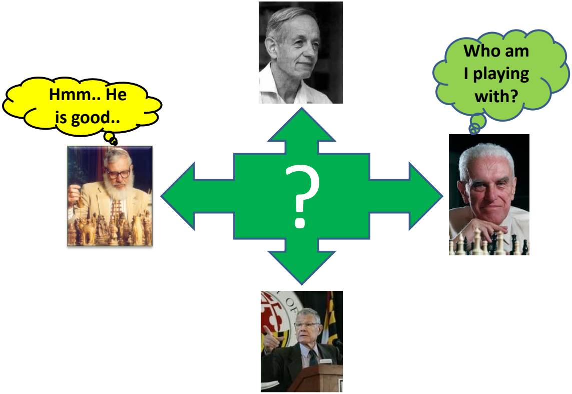
- Two aspects to applying game theory to engineering problems ...
  - **Utility design**
    - Selection of suitable operating points as **equilibria** of game
    - Desirable properties at equilibria – efficiency, fairness
  - **Algorithm design** or **(adaptive) dynamics** – **<u>Focus of this talk</u>**
    - Convergence to desired operating point
    - **Robust to feedback delays**
    - **Resilient to perturbation**

Overview

- Background
- **Setup**
- Classification of games
- Overview of existing literature
- Main results
  - Algorithm #1 – Generalized Better Reply Path Algorithm
  - Algorithm #2 – Simple Experimentation with Monitoring
- Future directions

Learning in Games (1)

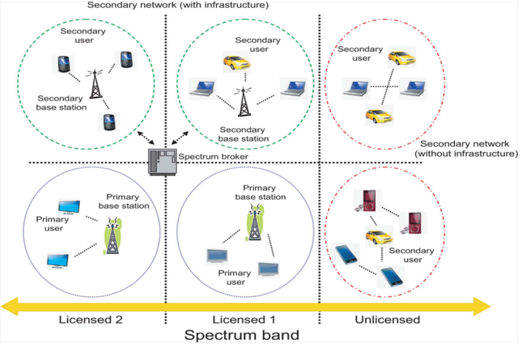
- **Incomplete-information (stage) game** – (some) agents unaware of the structure of the game
  - May not be aware of other agents
  - May not even be aware that they are playing games



Learning in Games (2)

- **Players interact with each other many times**
  - Can learn from the past payoffs and, possibly, actions of other players

- Examples: Dynamic channel access in cognitive radio, wireless sensor networks



Setup (1)

- Finite **stage game** (or **one-shot game**) in normal-form
  - $\mathcal{P} = \{1, 2, \ldots, n\}$ – set of n agents or players
  - **Pure action space**: $\mathcal{A}_i = \{1, 2, \ldots, A_i\}$ – set of A_i **pure actions** or **strategies** for agent $i \in \mathcal{P}$
  - **Payoff function**: $U_i : \mathcal{A} \to \mathbb{R}$
    - $U_i(\mathbf{a})$ is the payoff of agent i when action profile $\mathbf{a} = (a_1, \ldots, a_n) \in \mathcal{A}$ is played

- Terminology and notation
  - **Mixed strategy** of agent i: $\mathbf{p}_i \in \Delta(\mathcal{A}_i)$ – a probability distribution over pure action space $\mathcal{A}_i$
  - **Pure action/strategy profile**: $(a_1, a_2, \ldots, a_n) \in \mathcal{A} := \prod_{i \in \mathcal{P}} \mathcal{A}_i$
    - For $i \in \mathcal{P}$, $\mathbf{a}_{-i} \in \mathcal{A}_{-i} := \prod_{j \neq i} \mathcal{A}_j$
    - Given $J \subset \mathcal{P}$, $\mathbf{a}_J \in \mathcal{A}_J := \prod_{i \in J} \mathcal{A}_i$

Setup (2)

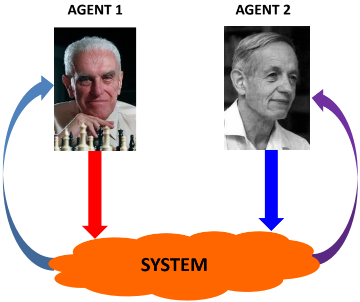
- **Pure-strategy Nash equilibrium** (PSNE) of stage game
  - Action profile $\mathbf{a}^\star = (a_1^\star, \ldots, a_n^\star) \in \mathcal{A}$ is a PSNE if, for all $i \in \mathcal{P}$,

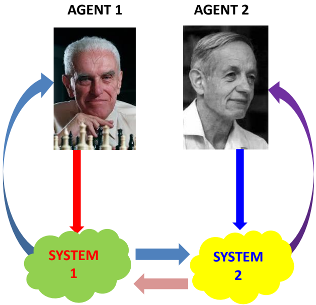
$$U_i(\mathbf{a}^\star) = \max_{a_i \in \mathcal{A}_i} U_i(a_i, \mathbf{a}^\star_{-i})$$

  - No agent can increase its own payoff through **unilateral deviation**
- Denote the set of PSNEs by $\mathcal{A}_{NE}$
  - We will assume $\mathcal{A}_{NE}$ is nonempty

Setup (3)

- Two different views of a game





- Global economy
- Markets
- Auctions

- Interconnected systems
- Regional economies

Setup (4)

- **Interactions among agents** over time modeled as (infinitely) **repeated game**
    - **Stage game** repeated at every $t \in \mathbb{N} := \{1, 2, \ldots\}$
        - Action profile selected at time t – $\mathbf{A}(t) = (A_i(t),\ i \in \mathcal{P})$
    - Agents update their (mixed) strategies via **learning rules**
- Focus on **uncoupled dynamics** – updates of an agent's action/strategy do **not** depend on the payoff functions of others
    - Players unaware of payoff functions of others (or even other players)

Setup (5)

- **Impossibility** result
  - "Uncoupled dynamics do not lead to Nash equilibrium," Hart and Mas-Colell, *The American Economic Review* (2003)

"There exists no uncoupled dynamics which guarantee Nash convergence"

- **Question of interest:** When does $\mathbf{A}(t)$ converge to an equilibrium (in an appropriate sense) as $t \to \infty$?

Overview

- Background
- Setup
- **Classification of games**
- Overview of existing literature
- Main results
  - Algorithm #1 – Generalized Better Reply Path Algorithm
  - Algorithm #2 – Simple Experimentation with Monitoring
- Future directions

Classification of games (1)

- **Identical interest games**
  - Payoff functions of all players are identical, i.e., there exists some function $\Phi : \mathcal{A} \to \mathbb{R}$ such that

$$U_i(\mathbf{a}) = \Phi(\mathbf{a}) \quad \text{for all} \quad i \in \mathcal{P} \quad \text{and} \quad \mathbf{a} \in \mathcal{A}$$

  - At least one PSNE
    - Maximizer of Φ

- **Potential games** (Rosenthal 1973)
  - There exists **potential function** $\Psi : \mathcal{A} \to \mathbb{R}$ such that, for all $i \in \mathcal{P}$, $\mathbf{a}_{-i} \in \mathcal{A}_{-i}$ and $a_i, a_i^* \in \mathcal{A}_i$,

$$U_i(a_i, \mathbf{a}_{-i}) - U_i(a_i^*, \mathbf{a}_{-i}) = \Psi(a_i, \mathbf{a}_{-i}) - \Psi(a_i^*, \mathbf{a}_{-i})$$

    - Change in an agent's payoff resulting from a unilateral change in action equal to the change in the "potential" function
  - At least one PSNE
    - Maximizer of potential function Ψ

Classification of games (2)

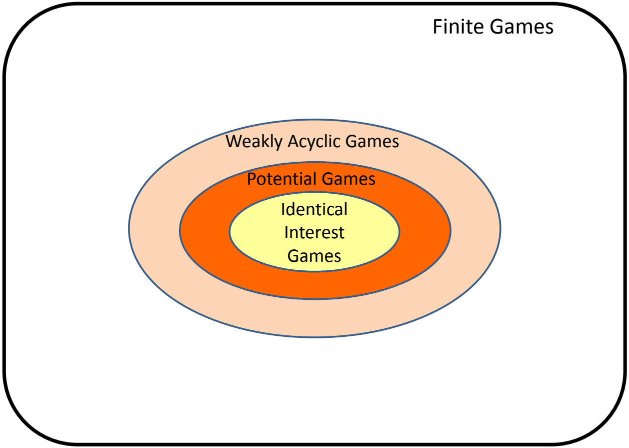
- **Weakly acyclic games** (Young 1993)
  - There exists a **global objective function** $\Omega : \mathcal{A} \to \mathbb{R}$ such that, for all $\mathbf{a} \in \mathcal{A}$ which is not a PSNE, there exist $i^* \in \mathcal{P}$ and $a_{i^*}^\dagger \in \mathcal{A}_{i^*}$ so that

    $$U_i(a_{i^*}^\dagger, \mathbf{a}_{-i^*}) > U_i(\mathbf{a}) \text{ and } \Omega(a_{i^*}^\dagger, \mathbf{a}_{-i^*}) > \Omega(\mathbf{a})$$

    - For any non-PSNE action profile, at least one agent's local payoff function is aligned with global objective function
  - **Alternate definition:** For every $\mathbf{a} \in \mathcal{A}$, there exists a **better reply path** $(\mathbf{a}(1), \ldots, \mathbf{a}(L))$ such that
    - $\mathbf{a}(1) = \mathbf{a}$ and $\mathbf{a}(L) \in \mathcal{A}_{NE}$
    - for all $\ell \in \{1, \ldots, L-1\}$, there is **exactly one agent** i^ℓ such that $a_{i^\ell}(\ell+1) \neq a_{i^\ell}(\ell)$ and $U_{i^\ell}(\mathbf{a}(\ell+1)) > U_{i^\ell}(\mathbf{a}(\ell))$

Learning in Games – Classification of games (3)

- Relation among different classes of games



Overview

- Background
- Setup
- Classification of games
- **Overview of existing literature**
- Main results
  - Algorithm #1 – Generalized Better Reply Path Algorithm
  - Algorithm #2 – Simple Experimentation with Monitoring
- Future directions

Existing literature on learning in games (1)

- **Fictitious play** (Brown 1951)
  - Players form beliefs about opponents' plays and behave rationally w.r.t. their **beliefs**

$$a_i(t) = \arg\max_{a_i \in \mathcal{A}_i} \sum_{\mathbf{a}_{-i} \in \mathcal{A}_{-i}} \mu_i^t(\mathbf{a}_{-i}) \cdot U_i(a_i, \mathbf{a}_{-i})$$

where

$$\mu_i^t(\mathbf{a}_{-i}) = \frac{n_i^t(\mathbf{a}_{-i})}{t-1} \quad \text{and} \quad n_i^t(\mathbf{a}_{-i}) = \sum_{\tau=1}^{t-1} \mathbf{1}\left\{\mathbf{A}_{-i}(\tau) = \mathbf{a}_{-i}\right\}$$

- **Regret matching** (Hart & Mas-Colell 2000)
  - At time $t+1 \in \mathbb{N}$, agent $i \in \mathcal{P}$ either
    - continues playing action $A_i(t) = a_i$, or
    - switches to other action $a_i^* \neq A_i(t)$ with probability proportional to **regret** $R_t^i(a_i, a_i^*)$ where

$$R_t^i(a_i, a_i^*) = \frac{1}{t}\left[\sum_{\tau \leq t : A_i(\tau) = a_i} \left(U_i(a_i^*, \mathbf{A}_{-i}(\tau)) - U_i(\mathbf{A}(\tau))\right)\right]^+$$

Existing literature on learning in games (2)

- **Regret testing** (Foster & Young 2003)

1. At time $t \in kT$, where $T > 1$ and $k \in \mathbb{N}$, each agent $i \in \mathcal{P}$ chooses a mixed strategy $\mathbf{p}_i(k) \in \Delta(\mathcal{A}_i)$
2. At time $t = kT, kT + 1, \ldots, (k+1)T - 1$, agent i chooses an action according to mixed strategy $\mathbf{p}_i(k)$
3. At time $t = (k+1)T$, agent i computes vector of **average regrets** over T periods

$$R^i_{a_i}(k) = \frac{1}{T} \sum_{\tau=kT}^{(k+1)T-1} \big(U_i(a_i, \mathbf{A}_{-i}(\tau)) - U_i(\mathbf{A}(\tau)) \big), \; a_i \in \mathcal{A}_i$$

4. If $R^i_{a_i}(k) \geq \rho$ $(\rho > 0)$ for some $a_i \in \mathcal{A}_i$, randomly choose a new mixed strategy $\mathbf{p}_i(k+1) \in \Delta(\mathcal{A}_i)$. Otherwise, $\mathbf{p}_i(k+1) = \mathbf{p}_i(k)$.
5. Increase k by one and go back to step 2

Existing literature on learning in games (3)

- Other learning rules
  - Efficient PSNE or socially efficient action profile – Pradelski and Young (2012), Marden, Young and Pao (2012), and Menon and Baras (2013)
  - Perfect foresight equilibrium
  - Many, many more!

Overview

- Setup
- Classification of game
- Background
- Overview of existing literature
- **Main results**
  - **Algorithm #1 – Generalized Better Reply Path Algorithm**
    - **Scenario #1 - Basic algorithm**
    - Scenario #2 - Feedback delays
    - Scenario #3 - Erroneous payoff estimates
    - Numerical examples
  - Algorithm #2 – Simple Experimentation with Monitoring
- Numerical examples
- Future directions

Basic algorithm (1)

- For every $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \mathcal{A}$ and $i \in \mathcal{P}$, define
$$BR_i(\mathbf{a}) = \{a_i^* \in \mathcal{A}_i \mid U_i(a_i^*, \mathbf{a}_{-i}) > U_i(\mathbf{a})\}$$
  - Set of **strictly better replies**

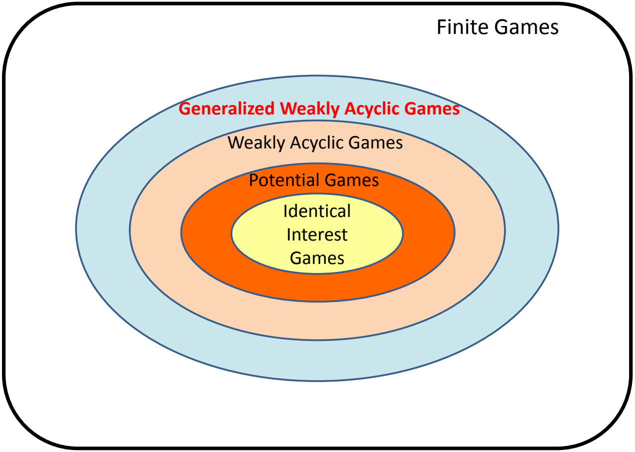
- **Generalized Better Reply Path Algorithm** (GBRPA)
  - At time $t = 2, 3, \ldots$, agent i chooses its action $a_i(t)$ as follows
    - If $BR_i(\mathbf{A}(t-1)) = \emptyset$
      $\diamond$ $A_i(t) = A_i(t-1)$
    - Else (i.e., $BR_i(\mathbf{A}(t-1)) \neq \emptyset$)
      $\diamond$ Choose $A_i(t) = a_i$ with probability
      $$\beta_i(a_i; \mathbf{A}(t-1)) \in [\underline{\epsilon}, \overline{\epsilon}]$$
      for all $a_i \in BR_i(\mathbf{A}(t-1))$, where $0 < \underline{\epsilon} \leq \overline{\epsilon} < 1$
      $\diamond$ Pick $A_i(t) = A_i(t-1)$ with prob.
      $$1 - \sum_{a_i \in BR_i(\mathbf{A}(t-1))} \beta_i(a_i; \mathbf{A}(t-1))$$

Basic algorithm (2)

- **Generalized weakly acyclic games** (Pal and La, ACC 2015)
  - **Generalized better reply path**: a sequence of action profiles $(\mathbf{a}(1), \ldots, \mathbf{a}(K))$, where for every $\ell = 1, \ldots, K-1$, there exists $\mathcal{I}(\ell) \subset \mathcal{P}$ such that
    - for all $i \in \mathcal{I}(\ell)$, $a_i(\ell) \neq a_i(\ell+1)$ and $U_i(\mathbf{a}(\ell)) < U_i(a_i(\ell+1), \mathbf{a}_{-i}(\ell))$
    - for all $i \neq \mathcal{I}(\ell)$, $a_i(\ell) = a_i(\ell+1)$
  - A game is **generalized weakly acyclic** if
    - $\mathcal{A}_{NE} \neq \emptyset$;
    - for all non-PSNE action profile $\mathbf{a} \in \mathcal{A} \setminus \mathcal{A}_{NE}$, there exists a **generalized better reply path** $(\mathbf{a}(1), \ldots, \mathbf{a}(L))$ with $\mathbf{a}(1) = \mathbf{a}$ and $\mathbf{a}(L) \in \mathcal{A}_{NE}$
  - Weakly acyclic games are special cases with $|\mathcal{I}(\ell)| = 1$

Basic algorithm (3)

- Relation among different classes of games



Basic algorithm (4)

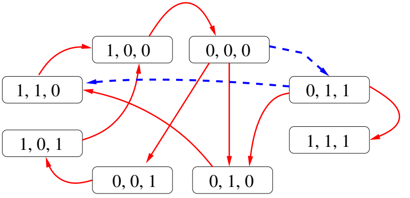
- Example of generalized weakly acyclic game that is **not** weakly acyclic
  - 3-player game with binary action space $\mathcal{A}_i = \{0, 1\}$, $i = 1, 2, 3$
  - Unique (weak) PSNE – (1, 1, 1)

		Player 2	
		0	1
Player 1	0	5, 5, 5	5, 7, 5
	1	4, 8, 5	6, 7, 5

$a_3 = 0$

		Player 2	
		0	1
Player 1	0	0, 0, 6	0, 0, 0
	1	1, 10, 0	10, 10, 5

$a_3 = 1$



- - - → simultaneous deviation by multiple players

——→ unilateral deviation

Basic algorithm (5)

Assumption

We assume $\max_{i \in \mathcal{P}} \left(\max_{\mathbf{a}^* \in \mathcal{A}} \sum_{a_i \in BR_i(\mathbf{a}^*)} \beta(a_i; \mathbf{a}^*) \right) < 1$

- Even when $BR_i(\mathbf{A}(t-1)) \neq \emptyset$, agent i chooses $A_i(t-1)$ at time t with *positive* probability

Theorem

Suppose that the game is **generalized weakly acyclic**. *Then, starting with any* **arbitrary** *initial action profile* $\mathbf{A}(1) = \mathbf{a} \in \mathcal{A}$*, the action profile converges to a PSNE almost surely under GBRPA. In other words,* **with probability 1** *(w.p.1), there exist finite* T^* *and a PSNE* $\mathbf{a}^\star$ *such that* $\mathbf{A}(t) = \mathbf{a}^\star$ *for all* $t \geq T^*$*.*

Basic algorithm (6)

Theorem

Suppose that the game is generalized weakly acyclic. Then, starting with an arbitrary initial action profile $\mathbf{A}(1) = \mathbf{a} \in \mathcal{A}$, *the probability* $\mathbb{P}\left[\mathbf{A}(t) \notin \mathcal{A}_{NE}\right]$ **decays geometrically** *under GBRPA, i.e., there exist* $C < \infty$ *and* $0 < \eta < 1$ *such that*

$$\mathbb{P}\left[\mathbf{A}(t) \notin \mathcal{A}_{NE}\right] \leq C \cdot \eta^t \quad \textit{for all} \quad t \in \mathbb{N}.$$

- Finite expected convergence time
- Parameter η depends on the **longest among the shortest generalized better reply paths** to a PSNE from non-PSNE action profiles

Basic algorithm (7)

Theorem

Suppose that the game is **not** *generalized weakly acyclic. Then, there exists at least one action profile* $\mathbf{a}^\star \in \mathcal{A}$ *such that, if* $\mathbf{A}(1) = \mathbf{a}^\star$, $\mathbf{A}(t) \notin \mathcal{A}_{NE}$ *for all* $t \in \mathbb{N}$.

- If $\mathbf{A}(1) \sim \mu$ and $\mu(\mathbf{a}) > 0$ for all $\mathbf{a} \in \mathcal{A}$, there is positive probability that the GBRPA will not converge to a PSNE ever
- GBRPA is guaranteed to converge to a PSNE, starting with any arbitrary initial action profile, **if and only if** the game is generalized weakly acyclic

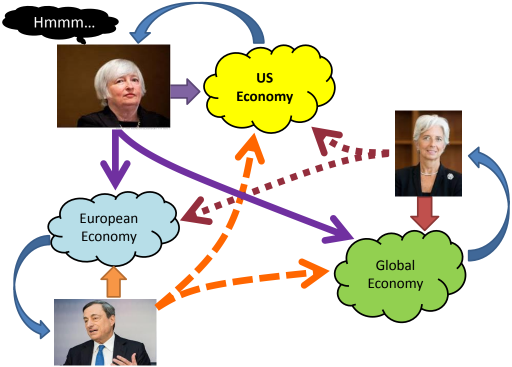
Overview

- Setup
- Classification of game
- Background
- Overview of existing literature
- **Main results**
  - **Algorithm #1 – Generalized Better Reply Path Algorithm**
    - Scenario #1 - Basic algorithm
    - **Scenario #2 - Feedback delays**
    - Scenario #3 - Erroneous payoff estimates
    - Numerical examples
  - Algorithm #2 – Simple Experimentation with Monitoring
- Numerical examples
- Future directions

Feedback delays (1)

- Delays in the system
  - **Forward delays** – delayed effects of new actions
  - **Feedback delays** – delayed realized payoff information

- **Example:** Economic policies implemented by various parties and their effects on the regional and global economies



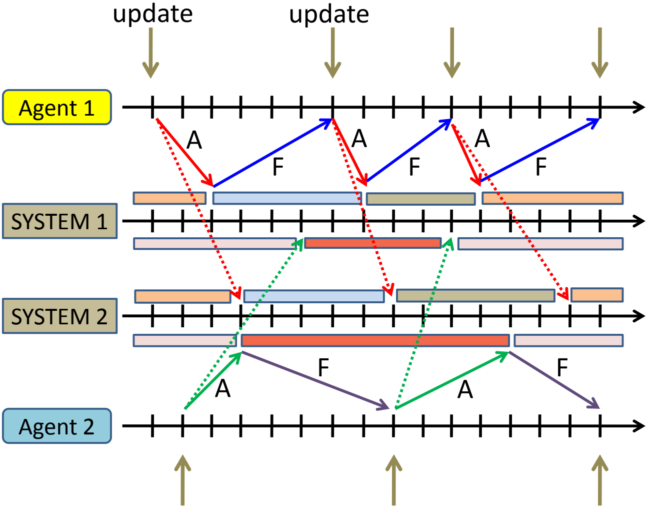
Feedback delays (2)

- Both **forward** and **feedback delays** experienced by agent $i \in \mathcal{P}$ modeled using **sequences of random variables**

- For the second view of a game
  - $\mathcal{T}^i = \{T_k^i,\ k \in \mathbb{Z}_+\}$, where T_k^i denotes the **time at which agent** i **updates its action** (or, equivalently, receives the payoff feedback) for the kth time with $T_0^i = 1$
    - $a_i(t) = a_i(T_k^i)$ for all $t \in \{T_k^i, \ldots, T_{k+1}^i - 1\}$, i.e., keeps the same action till next update
  - **Payoff (feedback) seen by agent** i at time T_k^i given by $U_i(\tilde{\mathbf{a}}^i(R_k^i))$, where $R_k^i \in \{T_{k-1}^i, \ldots, T_k^i - 1\}$
    - $\tilde{\mathbf{a}}^i(t)$ - action profile in effect at time t

Feedback delays (3)

- A picture is worth a thousand words ...



Feedback delays (4)

Theorem

Suppose that the game is **generalized weakly acyclic**. *Then, under some mild technical assumptions, starting with an* **arbitrary** *initial action profile* $\mathbf{A}(1) = \mathbf{a} \in \mathcal{A}$, *the action profile converges to a PSNE almost surely. In other words,* **w.p.1**, *there exist finite* T^* *and a PSNE* $\mathbf{a}^\star$ *such that* $\mathbf{A}(t) = \mathbf{a}^\star$ *for all* $t \geq T^*$.

- Delays have no effect on almost sure convergence of action profile to a PSNE under mild technical conditions

Overview

- Setup
- Classification of game
- Background
- Overview of existing literature
- **Main results**
    - Algorithm #1 - Generalized Better Reply Path Algorithm
        - Scenario #1 - Basic algorithm
        - Scenario #2 - Feedback delays
        - **Scenario #3 - Erroneous payoff estimates**
        - Numerical examples
    - Algorithm #2 - Simple Experimentation with Monitoring
- Numerical examples
- Future directions

Erroneous payoff estimation (1)

- In practice, agents may not be able to accurately determine $BR_i(\mathbf{A}(t))$
  - Noisy payoff measurements
- Agents may be able to determine them more reliably over time
- Let $p^i : \mathbb{N} \to [0,1]$, where $p^i(t)$ is the probability that agent i will **incorrectly** determine if action a_i belongs to $BR_i(\mathbf{a})$ at time t
  - Independent among actions

Assumption

There exists a decreasing, positive sequence $(\epsilon_t, t \in \mathbb{N})$ *such that*

i. $\lim_{t\to\infty} \epsilon_t = 0$, *and*

ii. *for every* $i \in \mathcal{P}$, *there are* $c_i > 0$ *and* $\gamma_i > 0$ *satisfying* $p^i(t) \sim c_i \cdot \epsilon_t^{\gamma_i}$.

Erroneous payoff estimation (2)

Theorem

Suppose that the game is **generalized weakly acyclic** *and* $\sum_{t \in \mathbb{N}} \epsilon_t^{\kappa} = \infty$, *where* κ *is a constant that satisfies some conditions. Then, under an additional mild technical condition,*

$$\lim_{t \to \infty} \mathbb{P}\left[\mathbf{A}(t) \in \mathcal{A}_{NE}\right] = 1.$$

- *Weaker* than almost sure convergence
- If $\epsilon_t \not\to 0$, but close to 0, then $\mathbf{A}(t) \in \mathcal{A}_{NE}$ with high probability for all sufficiently large t

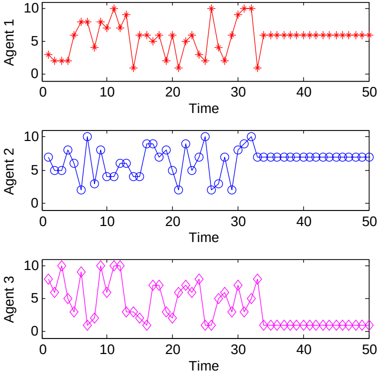
Overview

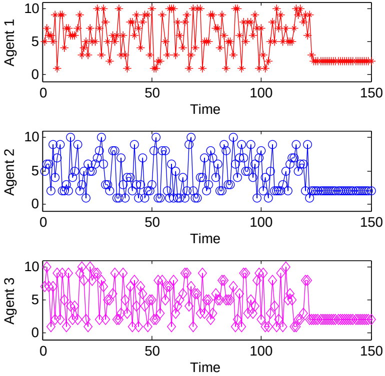
- Setup
- Classification of game
- Background
- Overview of existing literature
- **Main results**
  - **Algorithm #1 – Generalized Better Reply Path Algorithm**
    - Scenario #1 - Basic algorithm
    - Scenario #2 - Feedback delays
    - Scenario #3 - Erroneous payoff estimates
    - **Numerical examples**
  - Algorithm #2 – Simple Experimentation with Monitoring
- Future directions

Numerical example (1)

- 3 players with identical action space $\mathcal{A} = \{1, 2, \ldots, 10\}$
- Two PSNEs – (6, 7, 1) and (2, 2, 2)

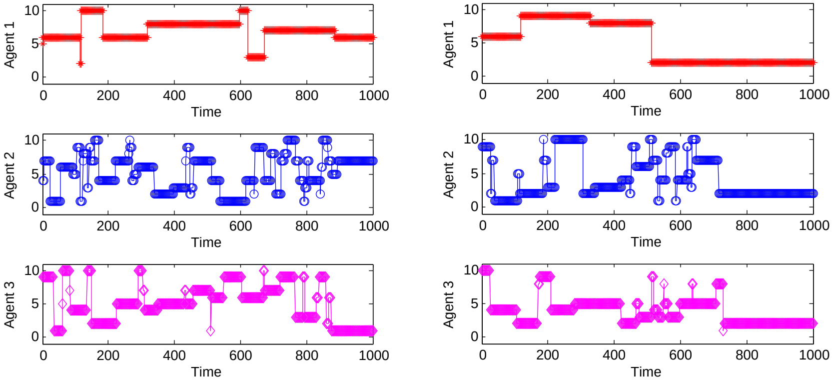
- No delays case





Numerical example (2)

- Forward delays $\sim$ geometric([0.01 0.1 0.05])
- Backward delays $= 1$



Overview

- Setup
- Classification of game
- Background
- Overview of existing literature
- **Main results**
  - Algorithm #1 - Generalized Better Reply Path Algorithm
  - **Algorithm #2 - Simple Experimentation with Monitoring**
- Future directions

Simple experimentation with monitoring (1)

- In practice,
  - Payoffs likely **noisy** or **random**

    > *"I regard this randomness as a crucial feature of many real-world games, where payoffs are likely to be affected by a wide assortment of forces that have been excluded when constructing the model"*
    >
    > – Larry Samuelson, *Evolutionary Games and Equilibrium Selection*

  - Agents may sometimes behave **irrationally**
    - **Faulty** or **unexpected** behavior

- **Question:** How do we select more **resilient** equilibrium?
  - Select **equilibria with a certain level of resilience**, or
  - Choose the **most resilient equilibria**

Simple experimentation with monitoring (2)

- **State** of an agent – (C)onverged, (E)xplore, a(L)ert
  - T alert states – $L_1, L_2, \ldots, L_T$
    - Still receiving the largest payoff possible, but on guard to determine if it needs to explore
  - State of agent $i \in \mathcal{P}$ at time $t \in \mathbb{N}$ denoted by $s_i(t)$

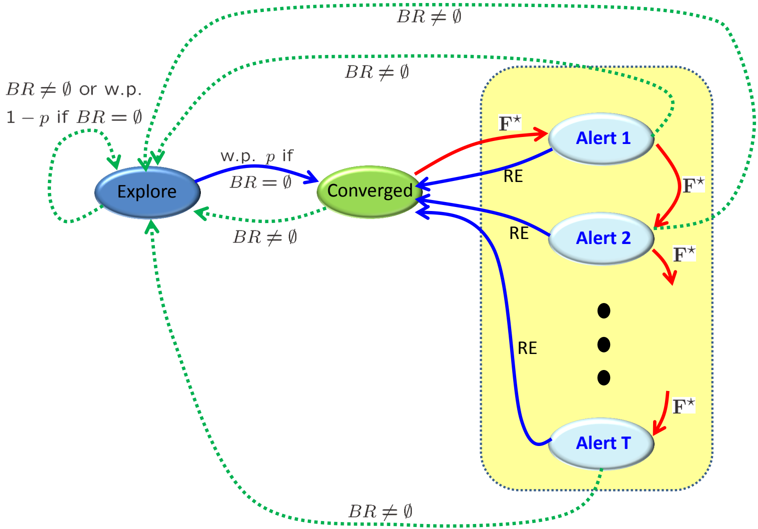
- **Algorithm #2** – **Simple Experimentation with Monitoring** (SEM)
  - **Action selection**
    - $s_i(t) = E \Longrightarrow \mathbb{P}\left[a_i(t) = a_i\right] \geq \delta > 0$ for all $a_i \in \mathcal{A}_i$
    - $s_i(t) = C$ or L_ℓ, $\ell = 1, 2, \ldots, T \Longrightarrow \mathbb{P}\left[a_i(t) = a_i(t-1)\right] = 1$
  - Occasional **faulty** or **irrational** behavior
    - At every $t \in \mathbb{N}$, each agent makes a mistake and chooses a **random** action with probability $\epsilon > 0$
    - Every action chosen with positive probability

Simple experimentation with monitoring (3)

- **State transition**
  - From **(C)**
    - If $BR_i(\mathbf{A}(t-1)) \neq \emptyset$, move to ($E$)
    - Elseif $BR_i(\mathbf{A}(t-1)) = \emptyset$ but the payoffs change (significantly), switch to (L_1)
      - Call this event $\mathbf{F}^\star$
    - Else, stay at (C)
  - From **(E)**
    - If $BR_i(\mathbf{A}(t-1)) \neq \emptyset$, stay at ($E$)
    - Else (i.e., $BR_i(\mathbf{A}(t-1)) = \emptyset$)
      - With prob. p ($0 < p < 1$), transition to (C)
      - With prob. $1 - p$, remain at (E)
  - From **(L_ℓ)**, $\ell = 1, \ldots, T$,
    - If $BR_i(\mathbf{A}(t-1)) \neq \emptyset$, move to ($E$)
    - Elseif the payoffs return to the expected payoffs last time at (C) (denoted RE), return to (C)
    - Else, jump to ($L_{\ell+1}$) if $\ell < T$ and (E) if $\ell = T$

Simple experimentation with monitoring (4)

- **State transitions**



Simple experimentation with monitoring (5)

- Define $d : \mathcal{A} \times \mathcal{A} \to \mathbb{Z}_+ := \{0, 1, 2, \ldots\}$, where

$$d(\mathbf{a}^1, \mathbf{a}^2) = \sum_{i \in \mathcal{P}} \mathbf{1}\left\{a_i^1 \neq a_i^2\right\}, \ \mathbf{a}^1, \mathbf{a}^2 \in \mathcal{A}$$

  - **Number of agents playing different actions**

- For $\tau \in \mathbb{Z}_+$, let $\mathcal{N}_\tau : \mathcal{A} \to 2^{\mathcal{A}}$, where

$$\mathcal{N}_\tau(\mathbf{a}) = \{\mathbf{a}' \mid d(\mathbf{a}, \mathbf{a}') \leq \tau\}, \ \mathbf{a} \in \mathcal{A}$$

- For each PSNE $\mathbf{a}^\star \in \mathcal{A}_{NE}$, define its **resilience** to be

$$R(\mathbf{a}^\star) = \max\{\tau \geq 0 \mid BR_i(a_i^\star, \mathbf{a}'_{-i}) = \emptyset \text{ for all } i \in \mathcal{P} \text{ and } \mathbf{a}' \in \mathcal{N}_\tau(\mathbf{a}^\star)\}$$

  - **Maximum number of deviations PSNE can tolerate** before unraveling

- The largest resilience among all PSNEs

$$R^\star_{\max} := \max_{\mathbf{a}^\star \in \mathcal{A}_{NE}} R(\mathbf{a}^\star)$$

Simple experimentation with monitoring (6)

Assumption

For all $\mathbf{a} \in \mathcal{A}$ *and for all* $J \subset \mathcal{P}$*, there exist (i)* $i \notin J$ *and (ii)* $\mathbf{a}^*_J \in \mathbf{A}_J$ *such that* $U_i(\mathbf{a}^*_J, \mathbf{a}_{-J}) \neq U_i(\mathbf{a})$ *(A4)*

- **Interdependence** assumption by Marden, Young and Pao (2012, IEEE CDC)

Simple experimentation with monitoring (7)

Theorem

Suppose that either Assumption (A4) or (A5) holds and $\mathcal{A}_{NE} \neq \emptyset$*. Then, one of the following holds as* $\epsilon \downarrow 0$*.*

- *If* $R^\star_{\max} < T$*, an action profile* $\mathbf{a} \in \mathcal{A}$ *is* **stochastically stable** *if and only if it is a PSNE and* $R(\mathbf{a}) = R^\star_{\max}$*.*
- *If* $R^\star_{\max} \geq T$*, an action profile* $\mathbf{a} \in \mathcal{A}$ *is* **stochastically stable** *if and only if it is a PSNE and* $R(\mathbf{a}) \geq T$*.*

- When ϵ is small, for all sufficiently large t, action profile $\mathbf{A}(t)$ lies in the set of stochastically stable PSNEs with high probability
- Allows us a means of choosing PSNEs with a certain level of resilience

Numerical example (1)

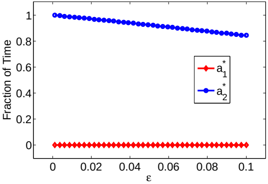
- 3 players with identical action space $\mathcal{A} = \{0, 1\}$
- Two PSNEs
  - $\mathbf{a}_1^\star = (0,\ 0,\ 0)$ – 0-resilient
  - $\mathbf{a}_2^\star = (1,\ 1,\ 1)$ – 1-resilient

		Player 2	
		0	1
Player 1	0	6,6,6	6,5,6
	1	5,6,6	10,10,6

a_3=0

		Player 2	
		0	1
Player 1	0	6,6,5	6,10,10
	1	10,6,10	10,10,10

a_3=1



Overview

- Setup
- Classification of game
- Background
- Overview of existing literature
- Main results
    - Algorithm #1 - Generalized Better Reply Path Algorithm
    - Algorithm #2 - Simple Experimentation with Monitoring
- **Future directions**

Future directions (1)

- Existence of global objective function for generalized weakly acyclic games
- Modeling random payoffs and examining their effects on algorithm design and resilience
- Joint utility and algorithm designs for efficiency and resilience

Acknowledgment

This work was in part funded by a grant from National Science Foundation (NSF) and grants from National Institute of Standards and Technology (NIST)