TECHNISCHE UNIVERSITÄT DRESDEN

Computational Logic ∴ Group

Hannes Strass

Faculty of Computer Science, Institute of Artificial Intelligence, Computational Logic Group

Counterfactual Regret Minimisation

Lecture 8, 10th Jun 2024 // Algorithmic Game Theory, SS 2024

Previously . . .

- A **behaviour strategy** assigns move probabilities to information sets.
- A **belief system** assigns probabilities to histories in information sets.
- An **assessment** is a pair (behaviour strategy profile, belief system).
- A **sequentially rational** assessment plays best responses “everywhere”.
- An assessment satisfies **consistency of beliefs** whenever the belief system’s probabilities match what is expected from everyone playing according to the behaviour strategy profile.
- An assessment is a **weak sequential equilibrium** iff it is both sequentially rational and satisfies consistency of beliefs.
- Mixed Nash equilibria for normal-form games and subgame perfect equilibria for sequential perfect-information games are special cases of weak sequential equilibria for extensive-form games.

Motivation

Main Question

- How to algorithmically solve imperfect-information games ...
- ... or at least devise good strategies or play them well in practice?

Motivation

Main Question

- How to algorithmically solve imperfect-information games ...
- ...or at least devise good strategies or play them well in practice?

Transformation to Normal Form?

Incurs an exponential blowup:

For every player $i \in P$, there are up to $|M_i|^{|\{ \mathcal{I}_j \in \mathcal{I} \mid p(\mathcal{I}_j) = i \}|}$ many behaviour strategies (pure strategies in the normal-form game).

Motivation

Main Question

- How to algorithmically solve imperfect-information games …
- …or at least devise good strategies or play them well in practice?

Transformation to Normal Form?

Incurs an exponential blowup:
For every player $i \in P$, there are up to $|M_i|^{|\{\mathcal{I}_j \in \mathcal{I} \mid p(\mathcal{I}_j)=i\}|}$ many behaviour strategies (pure strategies in the normal-form game).

Algorithms for sequential (perfect-information) games?

- Player i's best move in $\mathcal{I}_j \in \mathcal{I}$ depends on the player's beliefs $\beta_i : \mathcal{I}_j \to [0, 1]$.

Motivation

Main Question

- How to algorithmically solve imperfect-information games ...
- ...or at least devise good strategies or play them well in practice?

Transformation to Normal Form?

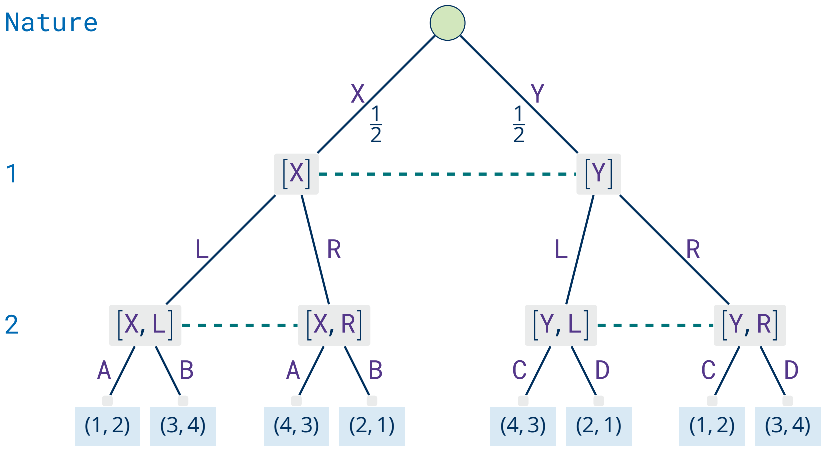
Incurs an exponential blowup:

For every player $i \in P$, there are up to $|M_i|^{|\{ \mathcal{I}_j \in \mathcal{I} \mid p(\mathcal{I}_j) = i \}|}$ many behaviour strategies (pure strategies in the normal-form game).

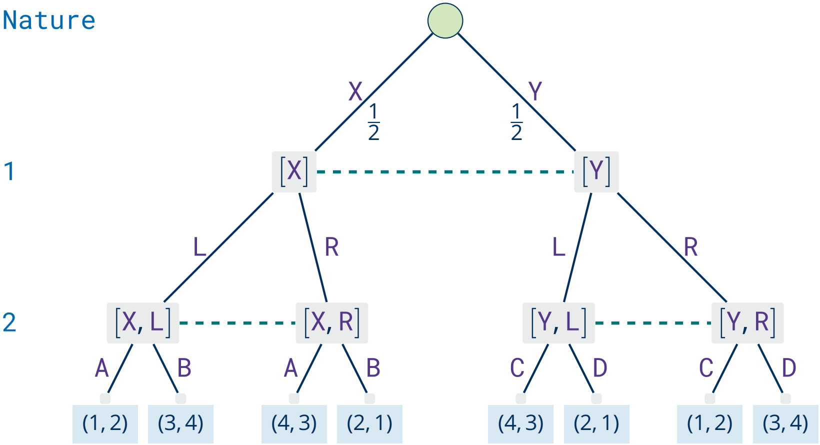
Algorithms for sequential (perfect-information) games?

- Player i's best move in $\mathcal{I}_j \in \mathcal{I}$ depends on the player's beliefs $\beta_i : \mathcal{I}_j \to [0, 1]$.
- Consistent beliefs about $\mathcal{I}_j$ in turn depend (in general) on probabilities of moves in other information sets (even on other paths of play).

Motivation: Example



Motivation: Example



The best move for 2 in {[X, L] , [X, R]} depends on what 2 does in {[Y, L] , [Y, R]}:

If 2 prefers C, then 1 will prefer L and thus 2 should prefer B. (Same for D and A.)

Motivation: Regret Matching

Before minimising regret in imperfect-information extensive-form games, we start with the simpler case of normal-form games . . .

Motivation: Regret Matching

Before minimising regret in imperfect-information extensive-form games, we start with the simpler case of normal-form games . . .

Recall

Let $(P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form, $i \in P$, and $s_j \in S_i$.

The **regret** of i playing s_j w.r.t. opponent profile $\boldsymbol{\pi}_{-i}$ is

$$r_{\boldsymbol{\pi}_{-i}, s_j} := \left(\max_{\pi_k \in \Pi_i} U_i(\boldsymbol{\pi}_{-i}, \pi_k) \right) - U_i(\boldsymbol{\pi}_{-i}, s_j)$$

Motivation: Regret Matching

Before minimising regret in imperfect-information extensive-form games, we start with the simpler case of normal-form games ...

Recall

Let $(P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form, $i \in P$, and $s_j \in S_i$.
The **regret** of i playing s_j w.r.t. opponent profile $\boldsymbol{\pi}_{-i}$ is

$$r_{\boldsymbol{\pi}_{-i}, s_j} := \left(\max_{\pi_k \in \Pi_i} U_i(\boldsymbol{\pi}_{-i}, \pi_k) \right) - U_i(\boldsymbol{\pi}_{-i}, s_j)$$

Difference between what player i could have had optimally vs. what they got.

Motivation: Regret Matching

Before minimising regret in imperfect-information extensive-form games, we start with the simpler case of normal-form games . . .

Recall

Let $(P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form, $i \in P$, and $s_j \in S_i$.
The **regret** of i playing s_j w.r.t. opponent profile $\boldsymbol{\pi}_{-i}$ is

$$r_{\boldsymbol{\pi}_{-i},s_j} := \left(\max_{\pi_k \in \Pi_i} U_i(\boldsymbol{\pi}_{-i}, \pi_k) \right) - U_i(\boldsymbol{\pi}_{-i}, s_j)$$

Difference between what player i could have had optimally vs. what they got.

Regret is zero iff a best response is played.

Motivation: Regret Matching

Before minimising regret in imperfect-information extensive-form games, we start with the simpler case of normal-form games . . .

Recall

Let $(P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form, $i \in P$, and $s_j \in S_i$.
The **regret** of i playing s_j w.r.t. opponent profile $\boldsymbol{\pi}_{-i}$ is

$$r_{\boldsymbol{\pi}_{-i},s_j} := \left(\max_{\pi_k \in \Pi_i} U_i(\boldsymbol{\pi}_{-i}, \pi_k) \right) - U_i(\boldsymbol{\pi}_{-i}, s_j)$$

Difference between what player i could have had optimally vs. what they got.

Regret is zero iff a best response is played.

⇝ Minimise regret over time in order to approach playing best responses.

Overview

Correlated Equilibria

Regret Matching

Counterfactual Regret Minimisation

Correlated Equilibria

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1,Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1,Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1, Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are (Stop, Go) and (Go, Stop):

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1,Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are (Stop, Go) and (Go, Stop):
  In both equilibria, one car never gets to move.

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1,Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are (Stop, Go) and (Go, Stop):
  In both equilibria, one car never gets to move.
- Another mixed Nash equilibrium is $\left(\left(\frac{100}{101}, \frac{1}{101}\right), \left(\frac{100}{101}, \frac{1}{101}\right)\right)$:

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1,Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are (Stop, Go) and (Go, Stop):
  In both equilibria, one car never gets to move.
- Another mixed Nash equilibrium is $\left(\left(\frac{100}{101}, \frac{1}{101}\right), \left(\frac{100}{101}, \frac{1}{101}\right)\right)$:
  Both cars mostly stop and there is a positive probability of accidents.

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1, Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are (Stop, Go) and (Go, Stop):
  In both equilibria, one car never gets to move.
- Another mixed Nash equilibrium is $\left(\left(\frac{100}{101}, \frac{1}{101}\right), \left(\frac{100}{101}, \frac{1}{101}\right)\right)$:
  Both cars mostly stop and there is a positive probability of accidents.
- A more desirable outcome would be: $\left\{(\texttt{Stop}, \texttt{Go}) \mapsto \frac{1}{2}, (\texttt{Go}, \texttt{Stop}) \mapsto \frac{1}{2}\right\}$:

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1, Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are (Stop, Go) and (Go, Stop):
  In both equilibria, one car never gets to move.
- Another mixed Nash equilibrium is $\left(\left(\frac{100}{101}, \frac{1}{101}\right), \left(\frac{100}{101}, \frac{1}{101}\right)\right)$:
  Both cars mostly stop and there is a positive probability of accidents.
- A more desirable outcome would be: $\left\{(\mathtt{Stop}, \mathtt{Go}) \mapsto \frac{1}{2}, (\mathtt{Go}, \mathtt{Stop}) \mapsto \frac{1}{2}\right\}$:
  However, mixed Nash equilibria cannot realise this.

Correlated Equilibria: Motivation

Traffic Lights

Two cars both want to cross an intersection. If a car stops, it does not get to the other side. If only one car goes, it gets to the other side. If both cars go, there is an accident.

(Car1, Car2)	Stop	Go
Stop	(0,0)	(0,1)
Go	(1,0)	(-100,-100)

- The pure Nash equilibria are (Stop, Go) and (Go, Stop):
  In both equilibria, one car never gets to move.
- Another mixed Nash equilibrium is $\left(\left(\frac{100}{101}, \frac{1}{101}\right), \left(\frac{100}{101}, \frac{1}{101}\right)\right)$:
  Both cars mostly stop and there is a positive probability of accidents.
- A more desirable outcome would be: $\left\{(\mathtt{Stop}, \mathtt{Go}) \mapsto \frac{1}{2}, (\mathtt{Go}, \mathtt{Stop}) \mapsto \frac{1}{2}\right\}$:
  However, mixed Nash equilibria cannot realise this. Traffic lights can!

Correlated Equilibrium: Intuition

- An external device chooses a strategy profile $\mathbf{s} \in \mathcal{S}$ randomly.

Correlated Equilibrium: Intuition

- An external device chooses a strategy profile $\mathbf{s} \in \mathcal{S}$ randomly.
- The distribution $\psi \colon \mathcal{S} \to [0, 1]$ for this is fixed and known to all players.

Correlated Equilibrium: Intuition

- An external device chooses a strategy profile $\mathbf{s} \in \mathcal{S}$ randomly.
- The distribution $\psi\colon \mathcal{S} \to [0, 1]$ for this is fixed and known to all players.
- For a chosen $(s_1, \ldots, s_n) \in \mathcal{S}$, each player $i \in P$ gets private advice $s_i \in S_i$.

Correlated Equilibrium: Intuition

- An external device chooses a strategy profile $\mathbf{s} \in \mathcal{S}$ randomly.
- The distribution $\psi \colon \mathcal{S} \to [0, 1]$ for this is fixed and known to all players.
- For a chosen $(s_1, \ldots, s_n) \in \mathcal{S}$, each player $i \in P$ gets private advice $s_i \in S_i$.
- Knowing $\{\mathbf{s} \in \mathcal{S} \mid \psi(\mathbf{s}) > 0\}$, player i may be able to infer advice of others.

Correlated Equilibrium: Intuition

- An external device chooses a strategy profile $\mathbf{s} \in \mathcal{S}$ randomly.
- The distribution $\psi\colon \mathcal{S} \to [0,1]$ for this is fixed and known to all players.
- For a chosen $(s_1, \ldots, s_n) \in \mathcal{S}$, each player $i \in P$ gets private advice $s_i \in S_i$.
- Knowing $\{\mathbf{s} \in \mathcal{S} \mid \psi(\mathbf{s}) > 0\}$, player i may be able to infer advice of others.
- Correlated equilibrium now means:
  No player has an incentive to deviate from the signal's advice.

Correlated Equilibrium: Intuition

- An external device chooses a strategy profile $\mathbf{s} \in \mathcal{S}$ randomly.
- The distribution $\psi\colon \mathcal{S} \to [0,1]$ for this is fixed and known to all players.
- For a chosen $(s_1, \ldots, s_n) \in \mathcal{S}$, each player $i \in P$ gets private advice $s_i \in S_i$.
- Knowing $\{\mathbf{s} \in \mathcal{S} \mid \psi(\mathbf{s}) > 0\}$, player i may be able to infer advice of others.
- Correlated equilibrium now means:
  No player has an incentive to deviate from the signal's advice.

Example

In the traffic lights game, assume $\psi = \left\{(\texttt{Stop}, \texttt{Go}) \mapsto \frac{1}{2}, (\texttt{Go}, \texttt{Stop}) \mapsto \frac{1}{2}, \ldots\right\}$:

- If `Car1` receives signal `Stop`, then it knows `Car2` must have received `Go`.
- Thus its best choice is to `Stop`.
- Symmetrically for `Car1` receiving signal `Go`, and `Car2`.

Correlated Equilibrium: Definition

Definition [Aumann, 1974]

Let $(P, \mathbf{S}, \mathbf{u})$ be a game in normal form with $P = \{1, \ldots, n\}$.

A probability distribution ψ on $\mathcal{S} = S_1 \times \ldots \times S_n$ is a **correlated equilibrium** iff for every $i \in P$, $s_j \in S_i$, and $s_k \in S_i$, we have

$$\sum_{\substack{\mathbf{s} \in \mathcal{S}, \\ \mathbf{s}_i = s_j}} \left(\psi(\mathbf{s}) \cdot \left(u_i(s_k, \mathbf{s}_{-i}) - u_i(\mathbf{s}) \right) \right) \leq 0$$

Roughly: Following the signal's advice incurs no (positive) regret.

Correlated Equilibrium: Definition

Definition [Aumann, 1974]

Let $(P, \mathbf{S}, \mathbf{u})$ be a game in normal form with $P = \{1, \ldots, n\}$.

A probability distribution ψ on $\mathcal{S} = S_1 \times \ldots \times S_n$ is a **correlated equilibrium** iff for every $i \in P$, $s_j \in S_i$, and $s_k \in S_i$, we have

$$\sum_{\substack{\mathbf{s} \in \mathcal{S}, \\ \mathbf{s}_i = s_j}} \left(\psi(\mathbf{s}) \cdot \left(u_i(s_k, \mathbf{s}_{-i}) - u_i(\mathbf{s}) \right) \right) \leq 0$$

Roughly: Following the signal's advice incurs no (positive) regret.

Observation

Every (mixed) Nash equilibrium $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ induces a correlated equilibrium $\psi_{\boldsymbol{\pi}} := \{(s_1, \ldots, s_n) \mapsto \pi_1(s_1) \cdot \ldots \cdot \pi_n(s_n) \mid (s_1, \ldots, s_n) \in \mathcal{S}\}$.

Correlated Equilibrium: Definition

Definition [Aumann, 1974]

Let $(P, \mathbf{S}, \mathbf{u})$ be a game in normal form with $P = \{1, \ldots, n\}$.

A probability distribution ψ on $\mathcal{S} = S_1 \times \ldots \times S_n$ is a **correlated equilibrium** iff for every $i \in P$, $s_j \in S_i$, and $s_k \in S_i$, we have

$$\sum_{\substack{\mathbf{s} \in \mathcal{S}, \\ \mathbf{s}_i = s_j}} \Big(\psi(\mathbf{s}) \cdot \big(u_i(s_k, \mathbf{s}_{-i}) - u_i(\mathbf{s}) \big) \Big) \leq 0$$

Roughly: Following the signal's advice incurs no (positive) regret.

Observation

Every (mixed) Nash equilibrium $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$ induces a correlated equilibrium $\psi_{\boldsymbol{\pi}} := \{(s_1, \ldots, s_n) \mapsto \pi_1(s_1) \cdot \ldots \cdot \pi_n(s_n) \mid (s_1, \ldots, s_n) \in \mathcal{S}\}$.

Correlated: Players no longer mix their strategies independently.

Correlated Equilibrium: Example (1)

Battle of the Partners

Two partners, Cat and Dee, think about how to spend the evening. Each has their personal preference what to do, but overall they want to spend the evening together.

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

Correlated Equilibrium: Example (1)

Battle of the Partners

Two partners, Cat and Dee, think about how to spend the evening. Each has their personal preference what to do, but overall they want to spend the evening together.

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

For the mixed Nash equilibrium $\boldsymbol{\pi} = (\pi_{\mathtt{Cat}}, \pi_{\mathtt{Dee}}) = \left(\left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{2}{3}\right)\right)$

Correlated Equilibrium: Example (1)

Battle of the Partners

Two partners, Cat and Dee, think about how to spend the evening. Each has their personal preference what to do, but overall they want to spend the evening together.

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

For the mixed Nash equilibrium $\boldsymbol{\pi} = (\pi_{\mathtt{Cat}}, \pi_{\mathtt{Dee}}) = \left(\left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{1}{3}, \frac{2}{3}\right)\right)$, we get

$$\psi_{\boldsymbol{\pi}} = \Big\{ (\mathtt{Cinema}, \mathtt{Cinema}) \mapsto \frac{2}{9}, (\mathtt{Cinema}, \mathtt{Dancing}) \mapsto \frac{4}{9}, (\mathtt{Dancing}, \mathtt{Cinema}) \mapsto \frac{1}{9}, (\mathtt{Dancing}, \mathtt{Dancing}) \mapsto \frac{2}{9} \Big\}$$

with $U_{\mathtt{Cat}}(\psi_{\boldsymbol{\pi}}) = U_{\mathtt{Dee}}(\psi_{\boldsymbol{\pi}}) = 4\frac{2}{3}$.

Correlated Equilibrium: Example (2)

To verify that $\psi_{\boldsymbol{\pi}}$ is a correlated equilibrium, we have the following cases:

Correlated Equilibrium: Example (2)

To verify that $\psi_{\boldsymbol{\pi}}$ is a correlated equilibrium, we have the following cases:

- $i = \texttt{Cat}, s_j = \texttt{Cinema}, s_k = \texttt{Dancing}$:

$$\psi(\texttt{Cinema},\texttt{Cinema}) \cdot (u_{\texttt{Cat}}(\texttt{Dancing},\texttt{Cinema}) - u_{\texttt{Cat}}(\texttt{Cinema},\texttt{Cinema})) + \psi(\texttt{Cinema},\texttt{Dancing}) \cdot (u_{\texttt{Cat}}(\texttt{Dancing},\texttt{Dancing}) - u_{\texttt{Cat}}(\texttt{Cinema},\texttt{Dancing})) = \frac{2}{9} \cdot (0 - 10) + \frac{4}{9} \cdot (7 - 2) = -\frac{20}{9} + \frac{20}{9} \leq 0$$

Correlated Equilibrium: Example (2)

To verify that $\psi_{\boldsymbol{\pi}}$ is a correlated equilibrium, we have the following cases:

- $i = \texttt{Cat}, s_j = \texttt{Cinema}, s_k = \texttt{Dancing}$:

$$\psi(\texttt{Cinema}, \texttt{Cinema}) \cdot (u_{\texttt{Cat}}(\texttt{Dancing}, \texttt{Cinema}) - u_{\texttt{Cat}}(\texttt{Cinema}, \texttt{Cinema})) + \psi(\texttt{Cinema}, \texttt{Dancing}) \cdot (u_{\texttt{Cat}}(\texttt{Dancing}, \texttt{Dancing}) - u_{\texttt{Cat}}(\texttt{Cinema}, \texttt{Dancing})) = \frac{2}{9} \cdot (0 - 10) + \frac{4}{9} \cdot (7 - 2) = -\frac{20}{9} + \frac{20}{9} \leq 0$$

- $i = \texttt{Cat}, s_j = \texttt{Dancing}, s_k = \texttt{Cinema}$:

$$\psi(\texttt{Dancing}, \texttt{Cinema}) \cdot (u_{\texttt{Cat}}(\texttt{Cinema}, \texttt{Cinema}) - u_{\texttt{Cat}}(\texttt{Dancing}, \texttt{Cinema})) + \psi(\texttt{Dancing}, \texttt{Dancing}) \cdot (u_{\texttt{Cat}}(\texttt{Cinema}, \texttt{Dancing}) - u_{\texttt{Cat}}(\texttt{Dancing}, \texttt{Dancing})) = \frac{1}{9} \cdot (10 - 0) + \frac{2}{9} \cdot (2 - 7) = \frac{10}{9} + \left(-\frac{10}{9}\right) \leq 0$$

Correlated Equilibrium: Example (2)

To verify that $\psi_{\boldsymbol{\pi}}$ is a correlated equilibrium, we have the following cases:

- $i = \mathtt{Cat}, s_j = \mathtt{Cinema}, s_k = \mathtt{Dancing}$:

$$\psi(\mathtt{Cinema}, \mathtt{Cinema}) \cdot (u_{\mathtt{Cat}}(\mathtt{Dancing}, \mathtt{Cinema}) - u_{\mathtt{Cat}}(\mathtt{Cinema}, \mathtt{Cinema})) +$$
$$\psi(\mathtt{Cinema}, \mathtt{Dancing}) \cdot (u_{\mathtt{Cat}}(\mathtt{Dancing}, \mathtt{Dancing}) - u_{\mathtt{Cat}}(\mathtt{Cinema}, \mathtt{Dancing})) =$$
$$\frac{2}{9} \cdot (0 - 10) + \frac{4}{9} \cdot (7 - 2) = -\frac{20}{9} + \frac{20}{9} \leq 0$$

- $i = \mathtt{Cat}, s_j = \mathtt{Dancing}, s_k = \mathtt{Cinema}$:

$$\psi(\mathtt{Dancing}, \mathtt{Cinema}) \cdot (u_{\mathtt{Cat}}(\mathtt{Cinema}, \mathtt{Cinema}) - u_{\mathtt{Cat}}(\mathtt{Dancing}, \mathtt{Cinema})) +$$
$$\psi(\mathtt{Dancing}, \mathtt{Dancing}) \cdot (u_{\mathtt{Cat}}(\mathtt{Cinema}, \mathtt{Dancing}) - u_{\mathtt{Cat}}(\mathtt{Dancing}, \mathtt{Dancing})) =$$
$$\frac{1}{9} \cdot (10 - 0) + \frac{2}{9} \cdot (2 - 7) = \frac{10}{9} + \left(-\frac{10}{9}\right) \leq 0$$

Due to $u_{\mathtt{Dee}}(s_1, s_2) = u_{\mathtt{Cat}}(s_2, s_1)$, this also covers the cases for $i = \mathtt{Dee}$.

Correlated Equilibria: Example (3)

Assume that both `Cat` and `Dee` have access to the result of one fair coin toss:

- If the coin shows heads, both go to the concert;
- if the coin shows tails, both go to the cinema.

Correlated Equilibria: Example (3)

Assume that both `Cat` and `Dee` have access to the result of one fair coin toss:

- If the coin shows heads, both go to the concert;
- if the coin shows tails, both go to the cinema.

This leads to the following (additional) correlated equilibrium:

$$\psi = \left\{(\mathtt{Cinema}, \mathtt{Cinema}) \mapsto \tfrac{1}{2}, (\mathtt{Dancing}, \mathtt{Dancing}) \mapsto \tfrac{1}{2}, \ldots\right\}$$

Correlated Equilibria: Example (3)

Assume that both `Cat` and `Dee` have access to the result of one fair coin toss:

- If the coin shows heads, both go to the concert;
- if the coin shows tails, both go to the cinema.

This leads to the following (additional) correlated equilibrium:

$$\psi = \left\{(\texttt{Cinema}, \texttt{Cinema}) \mapsto \tfrac{1}{2}, (\texttt{Dancing}, \texttt{Dancing}) \mapsto \tfrac{1}{2}, \ldots\right\}$$

with associated payoffs $U_{\texttt{Cat}}(\psi) = U_{\texttt{Dee}}(\psi) = \frac{1}{2} \cdot 10 + \frac{1}{2} \cdot 7 = 8\frac{1}{2}$.

Correlated Equilibria: Example (3)

Assume that both Cat and Dee have access to the result of one fair coin toss:

- If the coin shows heads, both go to the concert;
- if the coin shows tails, both go to the cinema.

This leads to the following (additional) correlated equilibrium:

$$\psi = \left\{(\mathtt{Cinema}, \mathtt{Cinema}) \mapsto \tfrac{1}{2}, (\mathtt{Dancing}, \mathtt{Dancing}) \mapsto \tfrac{1}{2}, \ldots\right\}$$

with associated payoffs $U_{\mathtt{Cat}}(\psi) = U_{\mathtt{Dee}}(\psi) = \frac{1}{2} \cdot 10 + \frac{1}{2} \cdot 7 = 8\frac{1}{2}$.

To verify that ψ is a correlated equilibrium, we (essentially) verify that:

$$\psi(\mathtt{Cinema}, \mathtt{Cinema}) \cdot (u_{\mathtt{Cat}}(\mathtt{Dancing}, \mathtt{Cinema}) - u_{\mathtt{Cat}}(\mathtt{Cinema}, \mathtt{Cinema})) \leq 0$$
$$\psi(\mathtt{Dancing}, \mathtt{Dancing}) \cdot (u_{\mathtt{Cat}}(\mathtt{Cinema}, \mathtt{Dancing}) - u_{\mathtt{Cat}}(\mathtt{Dancing}, \mathtt{Dancing})) \leq 0$$

which holds because $\frac{1}{2} \cdot (0 - 10) = -5 \leq 0$ and $\frac{1}{2} \cdot (2 - 7) = -2\frac{1}{2} \leq 0$.

Correlated Equilibria Form a Convex Set

Theorem

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a strategic game in normal form.

For any two correlated equilibria ψ_1 and ψ_2, and for any $\alpha \in [0, 1]$, we find that $\psi_\alpha := \{\mathbf{s} \mapsto \alpha \cdot \psi_1(\mathbf{s}) + (1 - \alpha) \cdot \psi_2(\mathbf{s}) \mid \mathbf{s} \in \mathcal{S}\}$ is a correlated equilibrium.

Correlated Equilibria Form a Convex Set

Theorem

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a strategic game in normal form.

For any two correlated equilibria ψ_1 and ψ_2, and for any $\alpha \in [0,1]$, we find that $\psi_\alpha := \{\mathbf{s} \mapsto \alpha \cdot \psi_1(\mathbf{s}) + (1-\alpha) \cdot \psi_2(\mathbf{s}) \mid \mathbf{s} \in \mathcal{S}\}$ is a correlated equilibrium.

Proof.

Let $\alpha \in [0,1]$ and consider any $i \in P$, $s_j, s_k \in S_i$. We have

$$\begin{aligned}
&\sum_{\mathbf{s}\in\mathcal{S},\mathbf{s}_i=s_j} \left(\psi_\alpha(\mathbf{s}) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))\right) \\
=& \sum_{\mathbf{s}\in\mathcal{S},\mathbf{s}_i=s_j} \left((\alpha \cdot \psi_1(\mathbf{s}) + (1-\alpha) \cdot \psi_2(\mathbf{s})) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))\right) \\
=& \sum_{\mathbf{s}\in\mathcal{S},\mathbf{s}_i=s_j} \Big((\alpha \cdot \psi_1(\mathbf{s}) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))) + ((1-\alpha) \cdot \psi_2(\mathbf{s})) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))\Big) \\
=& \sum_{\mathbf{s}\in\mathcal{S},\mathbf{s}_i=s_j} \Big(\alpha \cdot \psi_1(\mathbf{s}) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))\Big) + \sum_{\mathbf{s}\in\mathcal{S},\mathbf{s}_i=s_j} \Big((1-\alpha) \cdot \psi_2(\mathbf{s}) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))\Big) \\
=& \alpha \cdot \sum_{\mathbf{s}\in\mathcal{S},\mathbf{s}_i=s_j} \Big(\psi_1(\mathbf{s}) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))\Big) + (1-\alpha) \cdot \sum_{\mathbf{s}\in\mathcal{S},\mathbf{s}_i=s_j} \Big(\psi_2(\mathbf{s}) \cdot (u_i(s_k,\mathbf{s}_{-i}) - u_i(\mathbf{s}))\Big) \leq 0 \quad \square
\end{aligned}$$

Regret Matching

Learning to Play

Learning in Games: General Setting

- A (normal-form) game is played repeatedly for time points $t = 1, 2, \ldots$.
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^t$ played previously, and their payoffs to i.
- Using this information, the player can devise a (mixed) strategy π_i^{t+1} to play at time point $t + 1$.

Learning to Play

Learning in Games: General Setting

- A (normal-form) game is played repeatedly for time points $t = 1, 2, \ldots$.
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^t$ played previously, and their payoffs to i.
- Using this information, the player can devise a (mixed) strategy π_i^{t+1} to play at time point $t + 1$.

How can we evaluate whether a learner (player) is "doing well"?

Learning to Play

Learning in Games: General Setting

- A (normal-form) game is played repeatedly for time points $t = 1, 2, \ldots$.
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^t$ played previously, and their payoffs to i.
- Using this information, the player can devise a (mixed) strategy π_i^{t+1} to play at time point $t + 1$.

How can we evaluate whether a learner (player) is "doing well"?

Hindsight Rationality

After playing the game for $t \to \infty$ time points, the player "cannot think of" a function $\Phi \colon \Pi_i \to \Pi_i$ that would strictly increase their payoff in hindsight.

Learning to Play

Learning in Games: General Setting

- A (normal-form) game is played repeatedly for time points $t = 1, 2, \ldots$.
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^t$ played previously, and their payoffs to i.
- Using this information, the player can devise a (mixed) strategy π_i^{t+1} to play at time point $t + 1$.

How can we evaluate whether a learner (player) is "doing well"?

Hindsight Rationality

After playing the game for $t \to \infty$ time points, the player "cannot think of" a function $\Phi\colon \Pi_i \to \Pi_i$ that would strictly increase their payoff in hindsight.

Can learning (dynamic, local) lead to equilibria (static, global)?

Regret Matching

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \ldots, T$ and take the perspective of $i \in P$.

Regret Matching

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \ldots, T$ and take the perspective of $i \in P$.

At each time step $t \leq T$, i's one-time regret of not having played $s_k \in S_i$ is:

$$r_i^t(s_k) := u_i(s_k, \mathbf{s}_{-i}^t) - u_i(\mathbf{s}^t)$$

Regret Matching

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \ldots, T$ and take the perspective of $i \in P$.

At each time step $t \leq T$, i's one-time regret of not having played $s_k \in S_i$ is:

$$r_i^t(s_k) := u_i(s_k, \mathbf{s}_{-i}^t) - u_i(\mathbf{s}^t)$$

At time point T, the accumulated regret of a strategy $s_k \in S_i$ is thus:

$$R_i^T(s_k) := \sum_{1 \leq t \leq T} r_i^t(s_k)$$

Regret Matching

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \ldots, T$ and take the perspective of $i \in P$.

At each time step $t \leq T$, i's one-time regret of not having played $s_k \in S_i$ is:

$$r_i^t(s_k) := u_i(s_k, \mathbf{s}_{-i}^t) - u_i(\mathbf{s}^t)$$

At time point T, the accumulated regret of a strategy $s_k \in S_i$ is thus:

$$R_i^T(s_k) := \sum_{1 \leq t \leq T} r_i^t(s_k)$$

The probabilities at $T + 1$ are then set to be proportional to positive regret:

$$\pi_i^{T+1}(s_j) := \begin{cases} \dfrac{\left[R_i^T(s_j)\right]^+}{R_i^{T,+}} & \text{if } R_i^{T,+} > 0, \quad \text{where} \quad R_i^{T,+} := \displaystyle\sum_{s_k \in S_i} \left[R_i^T(s_k)\right]^+, \\ \dfrac{1}{|S_i|} & \text{otherwise.} \end{cases} \qquad \text{for } s_j \in S_i$$

$([x]^+ := \max\{x, 0\} \quad \text{for all } x \in \mathbb{R}.)$

Regret Matching

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \ldots, T$ and take the perspective of $i \in P$.

At each time step $t \leq T$, i's one-time regret of not having played $s_k \in S_i$ is:

$$r_i^t(s_k) := u_i(s_k, \mathbf{s}_{-i}^t) - u_i(\mathbf{s}^t)$$

At time point T, the accumulated regret of a strategy $s_k \in S_i$ is thus:

$$R_i^T(s_k) := \sum_{1 \leq t \leq T} r_i^t(s_k)$$

The probabilities at $T + 1$ are then set to be proportional to positive regret:

$$\pi_i^{T+1}(s_j) := \begin{cases} \dfrac{\left[R_i^T(s_j)\right]^+}{R_i^{T,+}} & \text{if } R_i^{T,+} > 0, \quad \text{where} \quad R_i^{T,+} := \sum_{s_k \in S_i} \left[R_i^T(s_k)\right]^+, \\ \dfrac{1}{|S_i|} & \text{otherwise.} \end{cases} \qquad \text{for } s_j \in S_i$$

$([x]^+ := \max\{x, 0\} \quad \text{for all } x \in \mathbb{R}.)$

Regret Matching: Example

$(\texttt{Cat}, \texttt{Dee})$	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $\overline{\texttt{Cinema}} = \texttt{Dancing}$ and $\overline{\texttt{Dancing}} = \texttt{Cinema}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$r^T_{\texttt{Cat}}\left(\overline{s^T_{\texttt{Cat}}}\right)$	$R^T_{\texttt{Cat}}(\texttt{Cinema})$	$R^T_{\texttt{Cat}}(\texttt{Dancing})$	$\pi^{T+1}_{\texttt{Cat}}$

Regret Matching: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $\overline{\texttt{Cinema}} = \texttt{Dancing}$ and $\overline{\texttt{Dancing}} = \texttt{Cinema}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$r^T_{\texttt{Cat}}\left(\overline{s^T_{\texttt{Cat}}}\right)$	$R^T_{\texttt{Cat}}(\texttt{Cinema})$	$R^T_{\texttt{Cat}}(\texttt{Dancing})$	$\pi^{T+1}_{\texttt{Cat}}$
1	(Cinema, Dancing)	5	0	5	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$

Regret Matching: Example

$(\texttt{Cat}, \texttt{Dee})$	$\texttt{Cinema}$	$\texttt{Dancing}$
$\texttt{Cinema}$	(10,7)	(2,2)
$\texttt{Dancing}$	(0,0)	(7,10)

We denote $\overline{\texttt{Cinema}} = \texttt{Dancing}$ and $\overline{\texttt{Dancing}} = \texttt{Cinema}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$r^T_{\texttt{Cat}}\left(\overline{s^T_{\texttt{Cat}}}\right)$	$R^T_{\texttt{Cat}}(\texttt{Cinema})$	$R^T_{\texttt{Cat}}(\texttt{Dancing})$	$\pi^{T+1}_{\texttt{Cat}}$
1	$(\texttt{Cinema}, \texttt{Dancing})$	5	0	5	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
2	$(\texttt{Dancing}, \texttt{Cinema})$	10	10	5	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$

Regret Matching: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $\overline{\text{Cinema}} = \text{Dancing}$ and $\overline{\text{Dancing}} = \text{Cinema}$.

T	$\mathbf{s}^T = \left(s^T_{\text{Cat}}, s^T_{\text{Dee}}\right)$	$r^T_{\text{Cat}}\left(\overline{s^T_{\text{Cat}}}\right)$	$R^T_{\text{Cat}}(\text{Cinema})$	$R^T_{\text{Cat}}(\text{Dancing})$	π^{T+1}_{Cat}
1	(Cinema, Dancing)	5	0	5	$\{\text{Cinema} \mapsto 0, \text{Dancing} \mapsto 1\}$
2	(Dancing, Cinema)	10	10	5	$\left\{\text{Cinema} \mapsto \frac{2}{3}, \text{Dancing} \mapsto \frac{1}{3}\right\}$
3	(Cinema, Dancing)	5	10	10	$\left\{\text{Cinema} \mapsto \frac{1}{2}, \text{Dancing} \mapsto \frac{1}{2}\right\}$

Regret Matching: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $\overline{\texttt{Cinema}} = \texttt{Dancing}$ and $\overline{\texttt{Dancing}} = \texttt{Cinema}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$r^T_{\texttt{Cat}}\left(\overline{s^T_{\texttt{Cat}}}\right)$	$R^T_{\texttt{Cat}}(\texttt{Cinema})$	$R^T_{\texttt{Cat}}(\texttt{Dancing})$	$\pi^{T+1}_{\texttt{Cat}}$
1	(Cinema, Dancing)	5	0	5	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
2	(Dancing, Cinema)	10	10	5	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$
3	(Cinema, Dancing)	5	10	10	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$
4	(Cinema, Cinema)	-10	10	0	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$

Regret Matching: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $\overline{\texttt{Cinema}} = \texttt{Dancing}$ and $\overline{\texttt{Dancing}} = \texttt{Cinema}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$r^T_{\texttt{Cat}}\left(\overline{s^T_{\texttt{Cat}}}\right)$	$R^T_{\texttt{Cat}}(\texttt{Cinema})$	$R^T_{\texttt{Cat}}(\texttt{Dancing})$	$\pi^{T+1}_{\texttt{Cat}}$
1	(Cinema, Dancing)	5	0	5	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
2	(Dancing, Cinema)	10	10	5	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$
3	(Cinema, Dancing)	5	10	10	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$
4	(Cinema, Cinema)	-10	10	0	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
5	(Cinema, Dancing)	5	10	5	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$

Regret Matching: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $\overline{\texttt{Cinema}} = \texttt{Dancing}$ and $\overline{\texttt{Dancing}} = \texttt{Cinema}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$r^T_{\texttt{Cat}}\left(\overline{s^T_{\texttt{Cat}}}\right)$	$R^T_{\texttt{Cat}}(\texttt{Cinema})$	$R^T_{\texttt{Cat}}(\texttt{Dancing})$	$\pi^{T+1}_{\texttt{Cat}}$
1	(Cinema, Dancing)	5	0	5	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
2	(Dancing, Cinema)	10	10	5	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$
3	(Cinema, Dancing)	5	10	10	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$
4	(Cinema, Cinema)	-10	10	0	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
5	(Cinema, Dancing)	5	10	5	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$
6	(Cinema, Cinema)	-10	10	-5	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$

Regret Matching: Correctness

For a given play sequence $(\mathbf{s}^t)_{t=1}^T$, and every $\mathbf{s}' \in \mathcal{S}$, define the **relative frequency** of $\mathbf{s}'$ after T rounds via

$$\bar{\varphi}^T(\mathbf{s}') := \frac{1}{T} \cdot \left|\left\{1 \leq t \leq T \;\middle|\; \mathbf{s}^t = \mathbf{s}'\right\}\right|$$

Theorem [Hart and Mas-Colell, 2000]

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form.

If every player plays according to regret matching, then $(\bar{\varphi}^t)_{t=1}^T$ converges to the set of correlated equilibria of G as $T \to \infty$.

Regret Matching: Correctness

For a given play sequence $(\mathbf{s}^t)_{t=1}^T$, and every $\mathbf{s}' \in \mathcal{S}$, define the **relative frequency** of $\mathbf{s}'$ after T rounds via

$$\bar{\varphi}^T(\mathbf{s}') := \frac{1}{T} \cdot \left| \left\{ 1 \leq t \leq T \;\middle|\; \mathbf{s}^t = \mathbf{s}' \right\} \right|$$

Theorem [Hart and Mas-Colell, 2000]

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form.

If every player plays according to regret matching, then $(\bar{\varphi}^t)_{t=1}^T$ converges to the set of correlated equilibria of G as $T \to \infty$.

More precisely: For any $\varepsilon > 0$, there is a $T_0 \geq 0$ such that for all $T > T_0$, there is a correlated equilibrium ψ_T of G whose distance from $\bar{\varphi}^T$ is at most ε.

Regret Matching: Correctness

For a given play sequence $(\mathbf{s}^t)_{t=1}^T$, and every $\mathbf{s}' \in \mathcal{S}$, define the **relative frequency** of $\mathbf{s}'$ after T rounds via

$$\bar{\varphi}^T(\mathbf{s}') := \frac{1}{T} \cdot \left|\left\{1 \leq t \leq T \;\middle|\; \mathbf{s}^t = \mathbf{s}'\right\}\right|$$

Theorem [Hart and Mas-Colell, 2000]

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form.
If every player plays according to regret matching, then $(\bar{\varphi}^t)_{t=1}^T$ converges to the set of correlated equilibria of G as $T \to \infty$.

More precisely: For any $\varepsilon > 0$, there is a $T_0 \geq 0$ such that for all $T > T_0$, there is a correlated equilibrium ψ_T of G whose distance from $\bar{\varphi}^T$ is at most ε.

Note: The result does not say that relative frequencies converge to a *point*.

Regret Matching: Correctness

For a given play sequence $(\mathbf{s}^t)_{t=1}^T$, and every $\mathbf{s}' \in \mathcal{S}$, define the **relative frequency** of $\mathbf{s}'$ after T rounds via

$$\bar{\varphi}^T(\mathbf{s}') := \frac{1}{T} \cdot \left|\left\{1 \leq t \leq T \;\middle|\; \mathbf{s}^t = \mathbf{s}'\right\}\right|$$

Theorem [Hart and Mas-Colell, 2000]

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a noncooperative game in normal form.
If every player plays according to regret matching, then $(\bar{\varphi}^t)_{t=1}^T$ converges to the set of correlated equilibria of G as $T \to \infty$.

More precisely: For any $\varepsilon > 0$, there is a $T_0 \geq 0$ such that for all $T > T_0$, there is a correlated equilibrium ψ_T of G whose distance from $\bar{\varphi}^T$ is at most ε.

Note: The result does not say that relative frequencies converge to a *point*.

$\rightsquigarrow$ Since all players must use regret matching, it will be used in self-play.

Regret Matching in Self-Play: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s_{\texttt{Cat}}^T, s_{\texttt{Dee}}^T\right)$	$R_{\texttt{Cat}}^T$	$R_{\texttt{Dee}}^T$	$\pi_{\texttt{Cat}}^{T+1}$	$\pi_{\texttt{Dee}}^{T+1}$

Regret Matching in Self-Play: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s_{\texttt{Cat}}^T, s_{\texttt{Dee}}^T\right)$	$R_{\texttt{Cat}}^T$	$R_{\texttt{Dee}}^T$	$\pi_{\texttt{Cat}}^{T+1}$	$\pi_{\texttt{Dee}}^{T+1}$
1	(Cinema, Dancing)	(0,5)	(5,0)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$

Regret Matching in Self-Play: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$R^T_{\texttt{Cat}}$	$R^T_{\texttt{Dee}}$	$\pi^{T+1}_{\texttt{Cat}}$	$\pi^{T+1}_{\texttt{Dee}}$
1	(Cinema, Dancing)	(0,5)	(5,0)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
2	(Dancing, Cinema)	(10,5)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$

Regret Matching in Self-Play: Example

$(\texttt{Cat}, \texttt{Dee})$	$\texttt{Cinema}$	$\texttt{Dancing}$
$\texttt{Cinema}$	(10,7)	(2,2)
$\texttt{Dancing}$	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s_{\texttt{Cat}}^T, s_{\texttt{Dee}}^T\right)$	$R_{\texttt{Cat}}^T$	$R_{\texttt{Dee}}^T$	$\pi_{\texttt{Cat}}^{T+1}$	$\pi_{\texttt{Dee}}^{T+1}$
1	$(\texttt{Cinema}, \texttt{Dancing})$	(0,5)	(5,0)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
2	$(\texttt{Dancing}, \texttt{Cinema})$	(10,5)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$
3	$(\texttt{Cinema}, \texttt{Dancing})$	(10,10)	(10,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$

Regret Matching in Self-Play: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s_{\texttt{Cat}}^T, s_{\texttt{Dee}}^T\right)$	$R_{\texttt{Cat}}^T$	$R_{\texttt{Dee}}^T$	$\pi_{\texttt{Cat}}^{T+1}$	$\pi_{\texttt{Dee}}^{T+1}$
1	(Cinema, Dancing)	(0,5)	(5,0)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
2	(Dancing, Cinema)	(10,5)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$
3	(Cinema, Dancing)	(10,10)	(10,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$
4	(Dancing, Dancing)	(5,10)	(0,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$

Regret Matching in Self-Play: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s_{\texttt{Cat}}^T, s_{\texttt{Dee}}^T\right)$	$R_{\texttt{Cat}}^T$	$R_{\texttt{Dee}}^T$	$\pi_{\texttt{Cat}}^{T+1}$	$\pi_{\texttt{Dee}}^{T+1}$
1	(Cinema, Dancing)	(0,5)	(5,0)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
2	(Dancing, Cinema)	(10,5)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$
3	(Cinema, Dancing)	(10,10)	(10,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$
4	(Dancing, Dancing)	(5,10)	(0,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
5	(Cinema, Dancing)	(5,15)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{4}, \texttt{Dancing} \mapsto \frac{3}{4}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$

Regret Matching in Self-Play: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s_{\texttt{Cat}}^T, s_{\texttt{Dee}}^T\right)$	$R_{\texttt{Cat}}^T$	$R_{\texttt{Dee}}^T$	$\pi_{\texttt{Cat}}^{T+1}$	$\pi_{\texttt{Dee}}^{T+1}$
1	(Cinema, Dancing)	(0,5)	(5,0)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
2	(Dancing, Cinema)	(10,5)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$
3	(Cinema, Dancing)	(10,10)	(10,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$
4	(Dancing, Dancing)	(5,10)	(0,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
5	(Cinema, Dancing)	(5,15)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{4}, \texttt{Dancing} \mapsto \frac{3}{4}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$
6	(Dancing, Dancing)	(0,15)	(-5,10)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$

Regret Matching in Self-Play: Example

(Cat, Dee)	Cinema	Dancing
Cinema	(10,7)	(2,2)
Dancing	(0,0)	(7,10)

We denote $R_i^T = \left(R_i^T(\texttt{Cinema}), R_i^T(\texttt{Dancing})\right)$ for $i \in \{\texttt{Cat}, \texttt{Dee}\}$.

T	$\mathbf{s}^T = \left(s^T_{\texttt{Cat}}, s^T_{\texttt{Dee}}\right)$	$R^T_{\texttt{Cat}}$	$R^T_{\texttt{Dee}}$	$\pi^{T+1}_{\texttt{Cat}}$	$\pi^{T+1}_{\texttt{Dee}}$
1	(Cinema, Dancing)	(0,5)	(5,0)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 1, \texttt{Dancing} \mapsto 0\}$
2	(Dancing, Cinema)	(10,5)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{2}{3}, \texttt{Dancing} \mapsto \frac{1}{3}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$
3	(Cinema, Dancing)	(10,10)	(10,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{2}, \texttt{Dancing} \mapsto \frac{1}{2}\right\}$
4	(Dancing, Dancing)	(5,10)	(0,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
5	(Cinema, Dancing)	(5,15)	(5,10)	$\left\{\texttt{Cinema} \mapsto \frac{1}{4}, \texttt{Dancing} \mapsto \frac{3}{4}\right\}$	$\left\{\texttt{Cinema} \mapsto \frac{1}{3}, \texttt{Dancing} \mapsto \frac{2}{3}\right\}$
6	(Dancing, Dancing)	(0,15)	(-5,10)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$
7	(Dancing, Dancing)	(-5,15)	(-15,10)	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$	$\{\texttt{Cinema} \mapsto 0, \texttt{Dancing} \mapsto 1\}$

Rate of Convergence

For a given sequence $(\boldsymbol{\pi}^t)_{t=1}^{T}$ of mixed-strategy profiles, define the (external) **overall regret** of player $i \in P$ after T rounds via

$$R_i^T := \max_{\hat{\pi} \in \Pi_i} \left\{ \sum_{t=1}^{T} \left(U_i(\hat{\pi}, \boldsymbol{\pi}_{-i}^t) - U_i(\pi_i^t, \boldsymbol{\pi}_{-i}^t) \right) \right\}$$

Rate of Convergence

For a given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles, define the (external) **overall regret** of player $i \in P$ after T rounds via

$$R_i^T := \max_{\hat{\pi} \in \Pi_i} \left\{ \sum_{t=1}^{T} \left(U_i(\hat{\pi}, \boldsymbol{\pi}_{-i}^t) - U_i(\pi_i^t, \boldsymbol{\pi}_{-i}^t) \right) \right\}$$

Theorem

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a normal-form game and let player $i \in P$ use regret matching in the sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles.

Then $R_i^T \leq \omega \cdot \sqrt{T}$, where the constant $\omega \in \mathbb{R}$ depends only on $\mathbf{u}$.

Rate of Convergence

For a given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles, define the (external) **overall regret** of player $i \in P$ after T rounds via

$$R_i^T := \max_{\hat{\pi} \in \Pi_i} \left\{ \sum_{t=1}^{T} \left(U_i(\hat{\pi}, \boldsymbol{\pi}_{-i}^t) - U_i(\pi_i^t, \boldsymbol{\pi}_{-i}^t) \right) \right\}$$

Theorem

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a normal-form game and let player $i \in P$ use regret matching in the sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles.

Then $R_i^T \leq \omega \cdot \sqrt{T}$, where the constant $\omega \in \mathbb{R}$ depends only on $\mathbf{u}$.

The **average overall regret** is then $\bar{R}_i^T := \frac{1}{T} \cdot R_i^T$.

Rate of Convergence

For a given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles, define the (external) **overall regret** of player $i \in P$ after T rounds via

$$R_i^T := \max_{\hat{\pi} \in \Pi_i} \left\{ \sum_{t=1}^{T} \left(U_i(\hat{\pi}, \boldsymbol{\pi}_{-i}^t) - U_i(\pi_i^t, \boldsymbol{\pi}_{-i}^t) \right) \right\}$$

Theorem

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a normal-form game and let player $i \in P$ use regret matching in the sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles.

Then $R_i^T \leq \omega \cdot \sqrt{T}$, where the constant $\omega \in \mathbb{R}$ depends only on $\mathbf{u}$.

The **average overall regret** is then $\bar{R}_i^T := \frac{1}{T} \cdot R_i^T$.

Proposition

$\bar{R}_i^T$ tends to zero as $T \to \infty$ iff $\bar{\varphi}^T$ tends to the set of correlated equilibria.

The Case of Two-Player Zero-Sum Games

For a given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles, define the **average mixed strategy** $\bar{\pi}_i^T$ of player $i \in P$ after T rounds via

$$\bar{\pi}_i^T(s_j) := \frac{1}{T} \cdot \sum_{t=1}^{T} \pi_i^t(s_j) \qquad \text{for } s_j \in S_i$$

The Case of Two-Player Zero-Sum Games

For a given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of mixed-strategy profiles, define the **average mixed strategy** $\bar{\pi}_i^T$ of player $i \in P$ after T rounds via

$$\bar{\pi}_i^T(s_j) := \frac{1}{T} \cdot \sum_{t=1}^{T} \pi_i^t(s_j) \qquad \text{for } s_j \in S_i$$

Theorem

Let $G = (P, \mathbf{S}, \mathbf{u})$ be a two-player, zero-sum normal-form game, i.e. $P = \{1, 2\}$, and let $(\boldsymbol{\pi}^t)_{t=1}^T$ be obtained from both players using regret matching.
Then as $T \to \infty$, the pair $\left(\bar{\pi}_1^T, \bar{\pi}_2^T\right)$ converges to the set of Nash equilibria of G.

Regret Matching$^+$

The computation of the accumulated (possibly negative) regret of a strategy $s_k \in S_i$ can be rewritten as:

$$R_i^T(s_k) := R_i^{T-1}(s_k) + r_i^T(s_k) \qquad \text{with } R_i^0(s_k) := 0$$

Regret Matching$^+$

The computation of the accumulated (possibly negative) regret of a strategy $s_k \in S_i$ can be rewritten as:

$$R_i^T(s_k) := R_i^{T-1}(s_k) + r_i^T(s_k) \qquad \text{with } R_i^0(s_k) := 0$$

Tammelin [2014] observed a better convergence when this is replaced by

$$R_i^{T,+}(s_k) := \left[R_i^{T-1}(s_k)\right]^+ + r_i^T(s_k)$$

Regret Matching$^+$

The computation of the accumulated (possibly negative) regret of a strategy $s_k \in S_i$ can be rewritten as:

$$R_i^T(s_k) := R_i^{T-1}(s_k) + r_i^T(s_k) \qquad \text{with } R_i^0(s_k) := 0$$

Tammelin [2014] observed a better convergence when this is replaced by

$$R_i^{T,+}(s_k) := \left[R_i^{T-1}(s_k)\right]^+ + r_i^T(s_k)$$

The probabilities at $T+1$ are again set to be proportional to positive regret:

$$\pi_i^{T+1}(s_j) := \begin{cases} \dfrac{R_i^{T,+}(s_j)}{R_i^{T,+}} & \text{if } R_i^{T,+} > 0, \quad \text{where} \quad R_i^{T,+} := \sum_{s_k \in S_i} R_i^{T,+}(s_k), \\ \dfrac{1}{|S_i|} & \text{otherwise.} \end{cases} \qquad \text{for } s_j \in S_i$$

Regret Matching$^+$

The computation of the accumulated (possibly negative) regret of a strategy $s_k \in S_i$ can be rewritten as:

$$R_i^T(s_k) := R_i^{T-1}(s_k) + r_i^T(s_k) \qquad \text{with } R_i^0(s_k) := 0$$

Tammelin [2014] observed a better convergence when this is replaced by

$$R_i^{T,+}(s_k) := \left[R_i^{T-1}(s_k)\right]^+ + r_i^T(s_k)$$

The probabilities at $T+1$ are again set to be proportional to positive regret:

$$\pi_i^{T+1}(s_j) := \begin{cases} \dfrac{R_i^{T,+}(s_j)}{R_i^{T,+}} & \text{if } R_i^{T,+} > 0, \quad \text{where} \quad R_i^{T,+} := \sum_{s_k \in S_i} R_i^{T,+}(s_k), \\ \dfrac{1}{|S_i|} & \text{otherwise.} \end{cases} \qquad \text{for } s_j \in S_i$$

RM$^+$ reacts more quickly when a previously poor action improves over time.

Counterfactual Regret Minimisation

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.
- Apply regret matching at each decision point (information set).

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.
- Apply regret matching at each decision point (information set).

Problem

Optimal moves depend on probabilities of moves in other information sets.

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.
- Apply regret matching at each decision point (information set).

Problem

Optimal moves depend on probabilities of moves in other information sets.

Solution of Zinkevich, Johanson, Bowling, and Piccione [2007]

- Define new notion of counterfactual regret:
  Assume the player played to deliberately reach a certain information set.

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.
- Apply regret matching at each decision point (information set).

Problem

Optimal moves depend on probabilities of moves in other information sets.

Solution of Zinkevich, Johanson, Bowling, and Piccione [2007]

- Define new notion of counterfactual regret:
  Assume the player played to deliberately reach a certain information set.
- Then for games with perfect recall:

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.
- Apply regret matching at each decision point (information set).

Problem

Optimal moves depend on probabilities of moves in other information sets.

Solution of Zinkevich, Johanson, Bowling, and Piccione [2007]

- Define new notion of counterfactual regret:

  Assume the player played to deliberately reach a certain information set.
- Then for games with perfect recall:
  - Regret matching can be applied to each information set independently.

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.
- Apply regret matching at each decision point (information set).

Problem

Optimal moves depend on probabilities of moves in other information sets.

Solution of Zinkevich, Johanson, Bowling, and Piccione [2007]

- Define new notion of counterfactual regret:

  Assume the player played to deliberately reach a certain information set.
- Then for games with perfect recall:
  - Regret matching can be applied to each information set independently.
  - Counterfactual regret is an upper bound for actual regret (main theorem).

From Normal Form to Extensive Form

Solving Imperfect-Information Games: Main Ideas

- Traverse the game tree in a backward induction-like fashion.
- Apply regret matching at each decision point (information set).

Problem

Optimal moves depend on probabilities of moves in other information sets.

Solution of Zinkevich, Johanson, Bowling, and Piccione [2007]

- Define new notion of counterfactual regret:

  Assume the player played to deliberately reach a certain information set.
- Then for games with perfect recall:
  - Regret matching can be applied to each information set independently.
  - Counterfactual regret is an upper bound for actual regret (main theorem).
  - Thus minimising counterfactual regret minimises actual regret.

Remember, Remember

Recall

$P(h' \mid h, \boldsymbol{\pi})$ is the probability that h' is reached when playing $\boldsymbol{\pi}$ from h on:

- $P(h \mid h, \boldsymbol{\pi}) = 1$ for all $h \in H$,
- $P([] \mid h, \boldsymbol{\pi}) = 0$ for all $h \neq []$, and
- $P([h'; m] \mid h, \boldsymbol{\pi}) = \pi_{p(\mathcal{I}_{h'})}(m \mid \mathcal{I}_{h'}) \cdot P(h' \mid h, \boldsymbol{\pi})$.

Remember, Remember

Recall

$P(h' \mid h, \boldsymbol{\pi})$ is the probability that h' is reached when playing $\boldsymbol{\pi}$ from h on:

- $P(h \mid h, \boldsymbol{\pi}) = 1$ for all $h \in H$,
- $P([] \mid h, \boldsymbol{\pi}) = 0$ for all $h \neq []$, and
- $P([h'; m] \mid h, \boldsymbol{\pi}) = \pi_{p(\mathcal{I}_{h'})}(m \mid \mathcal{I}_{h'}) \cdot P(h' \mid h, \boldsymbol{\pi})$.

Recall

The probability of reaching information set $\mathcal{I}_j$ when playing $\boldsymbol{\pi}$ is thus

$$P(\mathcal{I}_j \mid \boldsymbol{\pi}) := \sum_{h \in \mathcal{I}_j} P(h \mid \boldsymbol{\pi}) \qquad \text{where } P(h \mid \boldsymbol{\pi}) \text{ denotes } P(h \mid [], \boldsymbol{\pi})$$

Remember, Remember

Recall

$P(h' \mid h, \boldsymbol{\pi})$ is the probability that h' is reached when playing $\boldsymbol{\pi}$ from h on:

- $P(h \mid h, \boldsymbol{\pi}) = 1$ for all $h \in H$,
- $P([] \mid h, \boldsymbol{\pi}) = 0$ for all $h \neq []$, and
- $P([h'; m] \mid h, \boldsymbol{\pi}) = \pi_{p(\mathcal{I}_{h'})}(m \mid \mathcal{I}_{h'}) \cdot P(h' \mid h, \boldsymbol{\pi})$.

Recall

The probability of reaching information set $\mathcal{I}_j$ when playing $\boldsymbol{\pi}$ is thus

$$P(\mathcal{I}_j \mid \boldsymbol{\pi}) := \sum_{h \in \mathcal{I}_j} P(h \mid \boldsymbol{\pi}) \qquad \text{where } P(h \mid \boldsymbol{\pi}) \text{ denotes } P(h \mid [], \boldsymbol{\pi})$$

Recall

Player i's expected utility of playing $\boldsymbol{\pi}$ when history h has been reached is

$$U_i(\boldsymbol{\pi} \mid h) := \sum_{z \in Z} P(z \mid h, \boldsymbol{\pi}) \cdot u_i(z)$$

Towards Counterfactual Regret

Definition

Consider an extensive-form game with player $i \in P$ and information sets $\mathcal{I}$.

1. The **counterfactual probability** of playing to reach $h \in H$ is given by

$$P([] \mid \boldsymbol{\pi}_{-i}) = 1 \text{ and } P([h'; m] \mid \boldsymbol{\pi}_{-i}) := \begin{cases} \pi_k(m \mid h') \cdot P(h' \mid \boldsymbol{\pi}_{-i}) & \text{if } p(h') = k \neq i, \\ P(h' \mid \boldsymbol{\pi}_{-i}) & \text{otherwise.} \end{cases}$$

We counterfactually assume that i intentionally played to reach $\mathcal{I}_j$.

Towards Counterfactual Regret

Definition

Consider an extensive-form game with player $i \in P$ and information sets $\mathcal{I}$.

1. The **counterfactual probability** of playing to reach $h \in H$ is given by

$$P([\,] \mid \boldsymbol{\pi}_{-i}) = 1 \text{ and } P([h'; m] \mid \boldsymbol{\pi}_{-i}) := \begin{cases} \pi_k(m \mid h') \cdot P(h' \mid \boldsymbol{\pi}_{-i}) & \text{if } p(h') = k \neq i, \\ P(h' \mid \boldsymbol{\pi}_{-i}) & \text{otherwise.} \end{cases}$$

2. The **counterfactual probability** of playing to reach $\mathcal{I}_j \in \mathcal{I}$ is

$$P(\mathcal{I}_j \mid \boldsymbol{\pi}_{-i}) := \sum_{h \in \mathcal{I}_j} P(h \mid \boldsymbol{\pi}_{-i})$$

We counterfactually assume that i intentionally played to reach $\mathcal{I}_j$.

Towards Counterfactual Regret

Definition

Consider an extensive-form game with player $i \in P$ and information sets $\mathcal{I}$.

1. The **counterfactual probability** of playing to reach $h \in H$ is given by

$$P([]\,|\,\boldsymbol{\pi}_{-i}) = 1 \text{ and } P([h';m]\,|\,\boldsymbol{\pi}_{-i}) := \begin{cases} \pi_k(m\,|\,h') \cdot P(h'\,|\,\boldsymbol{\pi}_{-i}) & \text{if } p(h') = k \neq i, \\ P(h'\,|\,\boldsymbol{\pi}_{-i}) & \text{otherwise.} \end{cases}$$

2. The **counterfactual probability** of playing to reach $\mathcal{I}_j \in \mathcal{I}$ is

$$P(\mathcal{I}_j\,|\,\boldsymbol{\pi}_{-i}) := \sum_{h \in \mathcal{I}_j} P(h\,|\,\boldsymbol{\pi}_{-i})$$

3. The **counterfactual utility** of playing to reach $\mathcal{I}_j$ and then playing $\boldsymbol{\pi}$ is

$$U_i(\boldsymbol{\pi}\,|\,\mathcal{I}_j) = \frac{\sum_{h \in \mathcal{I}_j} P(h\,|\,\boldsymbol{\pi}_{-i}) \cdot U_i(\boldsymbol{\pi}\,|\,h)}{P(\mathcal{I}_j\,|\,\boldsymbol{\pi}_{-i})} = \frac{\sum_{h \in \mathcal{I}_j} P(h\,|\,\boldsymbol{\pi}_{-i}) \cdot \sum_{z \in Z} P(z\,|\,h, \boldsymbol{\pi}) \cdot u_i(z)}{P(\mathcal{I}_j\,|\,\boldsymbol{\pi}_{-i})}$$

We counterfactually assume that i intentionally played to reach $\mathcal{I}_j$.

Counterfactual Regret

Definition

Consider $i \in P$ and $\mathcal{I}_j \in \mathcal{I}$ with $p(\mathcal{I}_j) = i$.

1. Denote the set of legal moves of i in $\mathcal{I}_j$ by

$$M_i(\mathcal{I}_j) := \{m \in M_i \mid [h; m] \in H \text{ for some } h \in \mathcal{I}_j\}$$

Counterfactual Regret

Definition

Consider $i \in P$ and $\mathcal{I}_j \in \mathcal{I}$ with $p(\mathcal{I}_j) = i$.

1. Denote the set of legal moves of i in $\mathcal{I}_j$ by
$$M_i(\mathcal{I}_j) := \left\{ m \in M_i \;\middle|\; [h; m] \in H \text{ for some } h \in \mathcal{I}_j \right\}$$
2. For behaviour strategy profile $\boldsymbol{\pi}$ and move $m \in M_i(\mathcal{I}_j)$, define modified profile $\langle \boldsymbol{\pi} \rangle_m^{\mathcal{I}_j}$ to be just like $\boldsymbol{\pi}$, except that in $\mathcal{I}_j$ it always chooses m.

Counterfactual Regret

Definition

Consider $i \in P$ and $\mathcal{I}_j \in \mathcal{I}$ with $p(\mathcal{I}_j) = i$.

1. Denote the set of legal moves of i in $\mathcal{I}_j$ by
$$M_i(\mathcal{I}_j) := \left\{ m \in M_i \;\middle|\; [h; m] \in H \text{ for some } h \in \mathcal{I}_j \right\}$$
2. For behaviour strategy profile $\boldsymbol{\pi}$ and move $m \in M_i(\mathcal{I}_j)$, define modified profile $\langle \boldsymbol{\pi} \rangle_m^{\mathcal{I}_j}$ to be just like $\boldsymbol{\pi}$, except that in $\mathcal{I}_j$ it always chooses m.
3. The **immediate counterfactual regret** at time T is then defined by
$$r_i^T(\mathcal{I}_j) := \max_{m^* \in M_i(\mathcal{I}_j)} \sum_{t=1}^{T} P(\mathcal{I}_j \,|\, \boldsymbol{\pi}_{-i}^t) \cdot \left(U_i\left(\langle \boldsymbol{\pi}^t \rangle_{m^*}^{\mathcal{I}_j} \,\middle|\, \mathcal{I}_j \right) - U_i\left(\boldsymbol{\pi}^t \,\middle|\, \mathcal{I}_j \right) \right)$$
for any sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of behaviour strategy profiles.

Counterfactual Regret

Definition

Consider $i \in P$ and $\mathcal{I}_j \in \mathcal{I}$ with $p(\mathcal{I}_j) = i$.

1. Denote the set of legal moves of i in $\mathcal{I}_j$ by
$$M_i(\mathcal{I}_j) := \{m \in M_i \mid [h; m] \in H \text{ for some } h \in \mathcal{I}_j\}$$
2. For behaviour strategy profile $\boldsymbol{\pi}$ and move $m \in M_i(\mathcal{I}_j)$, define modified profile $\langle \boldsymbol{\pi} \rangle_m^{\mathcal{I}_j}$ to be just like $\boldsymbol{\pi}$, except that in $\mathcal{I}_j$ it always chooses m.
3. The **immediate counterfactual regret** at time T is then defined by
$$r_i^T(\mathcal{I}_j) := \max_{m^* \in M_i(\mathcal{I}_j)} \sum_{t=1}^{T} P(\mathcal{I}_j \mid \boldsymbol{\pi}_{-i}^t) \cdot \left(U_i\left(\langle \boldsymbol{\pi}^t \rangle_{m^*}^{\mathcal{I}_j} \,\middle|\, \mathcal{I}_j \right) - U_i\left(\boldsymbol{\pi}^t \mid \mathcal{I}_j\right) \right)$$
for any sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of behaviour strategy profiles.

Key Feature: r_i^T can be minimised by controlling only $\pi_i(\mathcal{I}_j)\colon M_i(\mathcal{I}_j) \to [0, 1]$.

Overall Regret $\leq$ Immediate Regret

Given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$, the (external) **overall regret** of player i at time T is:

$$R_i^T = \max_{\pi_i^* \in \Pi_i} \sum_{t=1}^{T} \left(U_i(\pi_i^*, \boldsymbol{\pi}_{-i}^t) - U_i(\boldsymbol{\pi}^t) \right)$$

where $U_i(\boldsymbol{\pi})$ denotes $U_i(\boldsymbol{\pi} \mid [])$.

Overall Regret $\leq$ Immediate Regret

Given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$, the (external) **overall regret** of player i at time T is:

$$R_i^T = \max_{\pi_i^* \in \Pi_i} \sum_{t=1}^{T} \left(U_i(\pi_i^*, \boldsymbol{\pi}_{-i}^t) - U_i(\boldsymbol{\pi}^t) \right)$$

where $U_i(\boldsymbol{\pi})$ denotes $U_i(\boldsymbol{\pi} \mid [\,])$.

Theorem [Zinkevich, Johanson, Bowling, and Piccione, 2007]

In any extensive-form game with perfect recall, for any player $i \in P$ and any sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of behaviour strategy profiles:

$$R_i^T \leq \sum_{\substack{\mathcal{I}_j \in \mathcal{I}, \\ p(\mathcal{I}_j) = i}} \left[r_i^T(\mathcal{I}_j) \right]^+$$

Overall Regret $\leq$ Immediate Regret

Given sequence $(\boldsymbol{\pi}^t)_{t=1}^T$, the (external) **overall regret** of player i at time T is:

$$R_i^T = \max_{\pi_i^* \in \Pi_i} \sum_{t=1}^{T} \left(U_i(\pi_i^*, \boldsymbol{\pi}_{-i}^t) - U_i(\boldsymbol{\pi}^t) \right)$$

where $U_i(\boldsymbol{\pi})$ denotes $U_i(\boldsymbol{\pi} \mid [])$.

Theorem [Zinkevich, Johanson, Bowling, and Piccione, 2007]

In any extensive-form game with perfect recall, for any player $i \in P$ and any sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of behaviour strategy profiles:

$$R_i^T \leq \sum_{\substack{\mathcal{I}_j \in \mathcal{I}, \\ p(\mathcal{I}_j) = i}} \left[r_i^T(\mathcal{I}_j) \right]^+$$

Thus: Minimising immediate regret in each $\mathcal{I}_j$ minimises overall regret.

Regret Matching at Information Sets

Definition

Consider the sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of behaviour strategy profiles of past play.

1. Let $\mathcal{I}_j \in \mathcal{I}$ with $p(\mathcal{I}_j) = i$ and $m \in M_i(\mathcal{I}_j)$. The **accumulated regret** of m is

$$R_i^T(\mathcal{I}_j, m) := \sum_{t=1}^{T} P(\mathcal{I}_j \mid \boldsymbol{\pi}_{-i}^t) \cdot \left(U_i\left(\langle \boldsymbol{\pi}^t \rangle_m^{\mathcal{I}_j} \,\middle|\, \mathcal{I}_j \right) - U_i\left(\boldsymbol{\pi}^t \,\middle|\, \mathcal{I}_j \right) \right)$$

Regret Matching at Information Sets

Definition

Consider the sequence $(\boldsymbol{\pi}^t)_{t=1}^T$ of behaviour strategy profiles of past play.

1. Let $\mathcal{I}_j \in \mathcal{I}$ with $p(\mathcal{I}_j) = i$ and $m \in M_i(\mathcal{I}_j)$. The **accumulated regret** of m is

$$R_i^T(\mathcal{I}_j, m) := \sum_{t=1}^{T} P(\mathcal{I}_j \mid \boldsymbol{\pi}_{-i}^t) \cdot \left(U_i\left(\langle \boldsymbol{\pi}^t \rangle_m^{\mathcal{I}_j} \,\middle|\, \mathcal{I}_j \right) - U_i\left(\boldsymbol{\pi}^t \,\middle|\, \mathcal{I}_j \right) \right)$$

2. The probability of playing m at $\mathcal{I}_j$ at time $T+1$ is set to

$$\pi_i^{T+1}(\mathcal{I}_j)(m) := \begin{cases} \dfrac{\left[R_i^T(\mathcal{I}_j, m)\right]^+}{R_i^{T,+}} & \text{if } R_i^{T,+} > 0, \quad \text{where } R_i^{T,+} := \sum_{m \in M_i(\mathcal{I}_j)} \left[R_i^T(\mathcal{I}_j, m)\right]^+ \\ \dfrac{1}{\left|M_i(\mathcal{I}_j)\right|} & \text{otherwise.} \end{cases}$$

CFR: Algorithm (1)

Initialisation of global variables:

function init() {
  foreach $i \in \{1, 2\}$ **do** {
    foreach $\mathcal{I}_j \in \mathcal{I}$ **with** $p(\mathcal{I}_j) = i$ **do** {
      foreach $m \in M_i(\mathcal{I}_j)$ **do** {
        $regret[j][m] := 0$ // accumulated regret table
        $strategy[j][m] := 0$ // accumulated strategy table
        $profile[1][j][m] := 1/\left|M_i(\mathcal{I}_j)\right|$ // move distribution for $\mathcal{I}_j$ at $t = 1$
      } } } }

CFR: Algorithm (1)

Initialisation of global variables:

```
function init() {
    foreach i ∈ {1,2} do {
        foreach 𝒥_j ∈ 𝒥 with p(𝒥_j) = i do {
            foreach m ∈ M_i(𝒥_j) do {
                regret[j][m] := 0                          // accumulated regret table
                strategy[j][m] := 0                        // accumulated strategy table
                profile[1][j][m] := 1/|M_i(𝒥_j)|           // move distribution for 𝒥_j at t = 1
            }}}}
```

Main Loop:

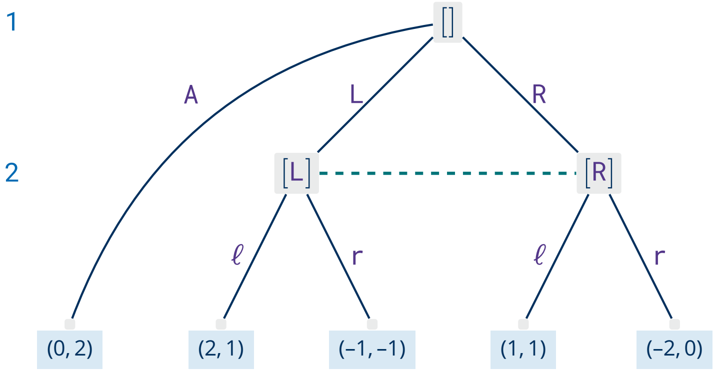
```
function solve(T) {
    foreach t ∈ {1,2,...,T} do {
        foreach i ∈ {1,2} do {
            cfr([], i, t, 1, 1)
        }}}
```

CFR: Algorithm (2)

```
function cfr(h, i, t, p_1, p_2) {                      // history, player, time point, reach probabilities
    if IS-TERMINAL(h) then return UTILITY_i(s)
    v_h := 0                                           // initialise expected payoff at h ∈ 𝓘_j
    foreach m ∈ M_p(𝓘_j)(𝓘_j) do { v'_h[j][m] := 0 }   // initialise payoffs of single moves
    foreach m ∈ M_p(𝓘_j)(𝓘_j) do {
        if TURN(h) = 1 then { v'_h[j][m] := cfr([h; m], i, t, profile[t][j][m] · p_1, p_2) }
            else { v'_h[j][m] := cfr([h; m], i, t, p_1, profile[t][j][m] · p_2) }
        v_h := v_h + profile[t][j][m] · v'_h[j][m] }   // accumulate currently expected payoff
    if TURN(h) = i then {                              // players minimise immediate regret of own moves
        r+ := 0                                        // initialise sum of positive regrets
        for m ∈ M_i(𝓘_j) do {                          // update values needed for regret matching
            regret[j][m] := regret[j][m] + p_{3−i} · (v'_h[m] − v_h)   // update accumulated cf regret
            strategy[j][m] := strategy[j][m] + p_i · profile[t][j][m]  // update “frequency” of move
            r+ := r+ + [regret[j][m]]+ }               // accumulate positive regret sum for normalisation
        if r+ > 0 then { foreach m ∈ M_i(𝓘_j) do {     // apply regret matching at 𝓘_j
                profile[t + 1][j][m] := [regret[j][m]]+ / r+ } }
            else { foreach m ∈ M_i(𝓘_j) do {
                profile[t + 1][j][m] := 1/|M_i(𝓘_j)| } } }
    return v_h }
```

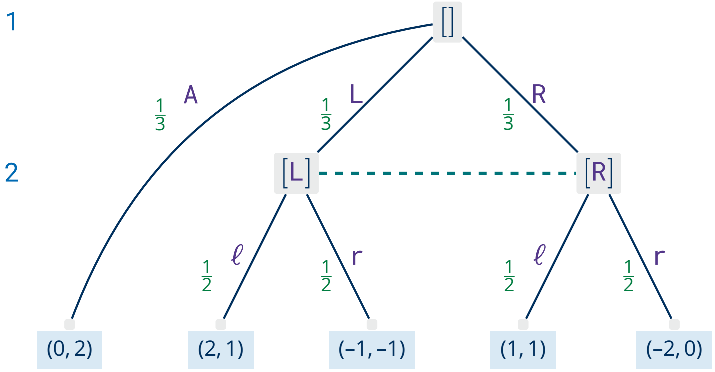
CFR: Example

Recall the following extensive-form game G_4:



CFR: Example

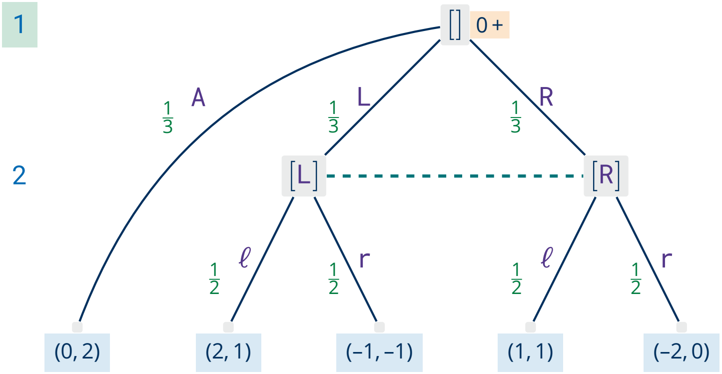
Recall the following extensive-form game G_4:



(1) Initialise move probabilities by uniform distributions

CFR: Example

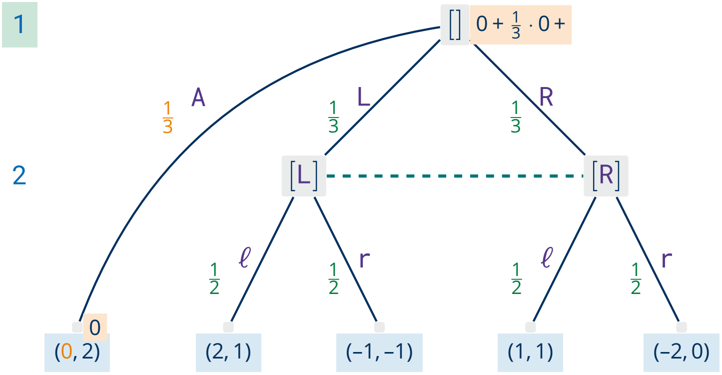
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

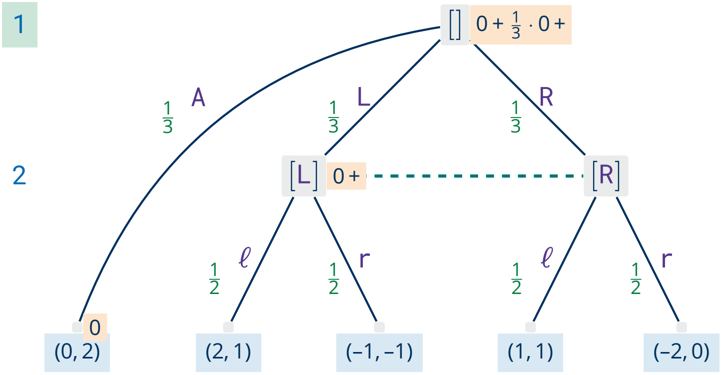
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

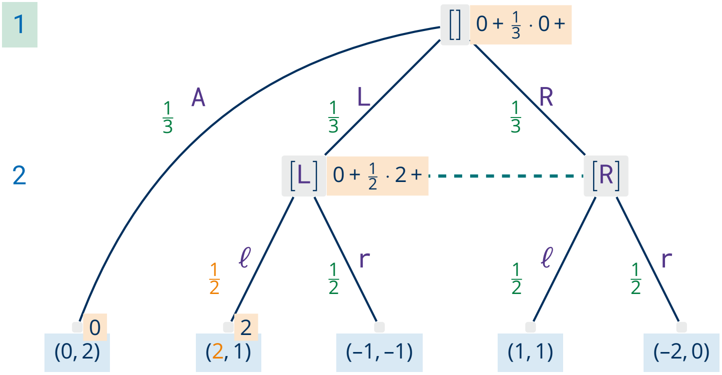
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

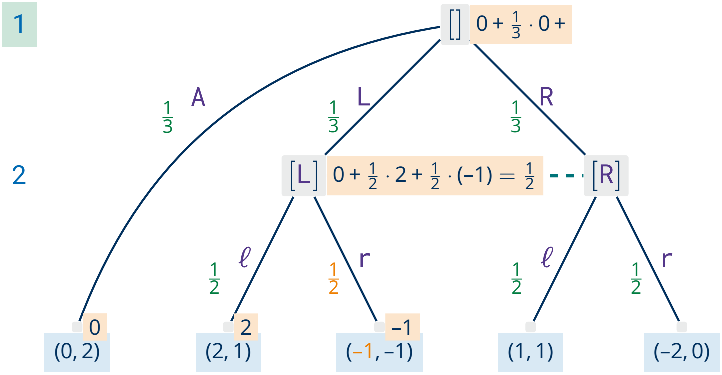
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1$, $i = 1$

CFR: Example

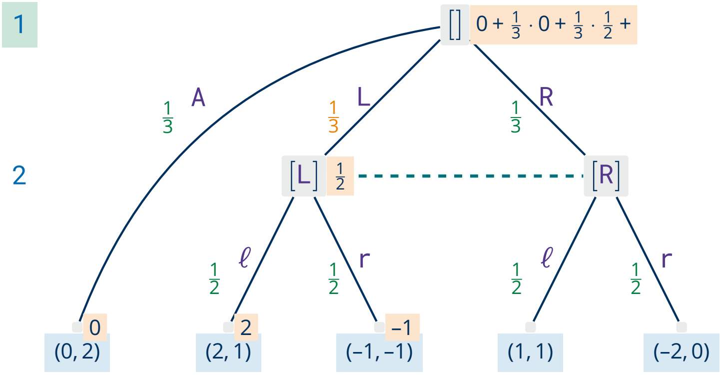
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

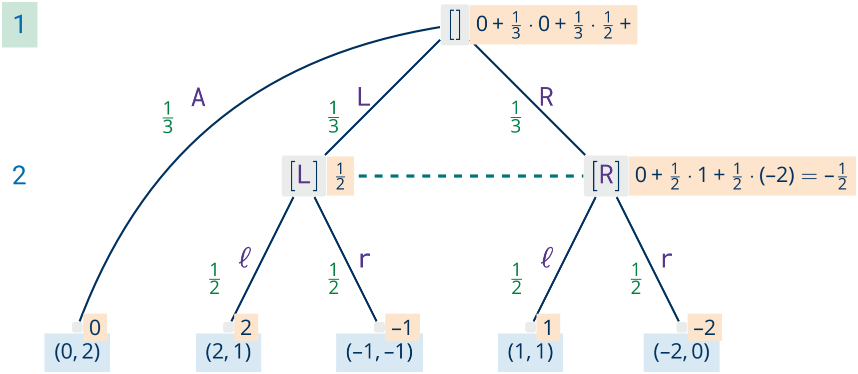
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

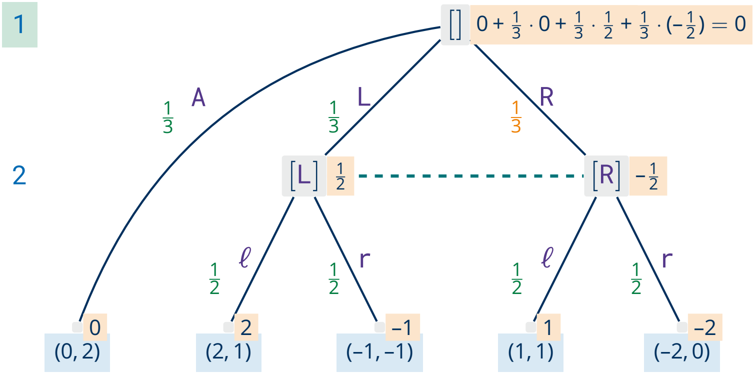
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

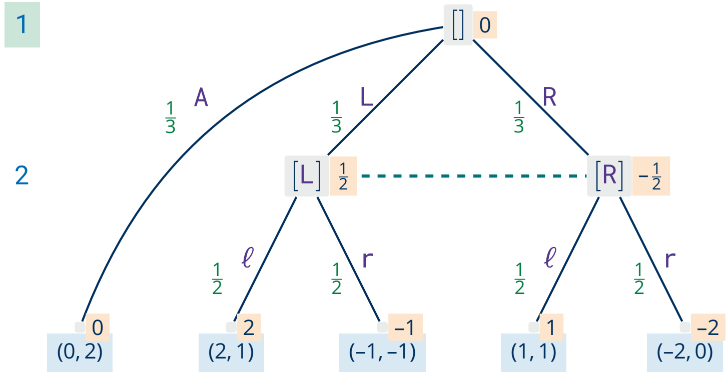
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

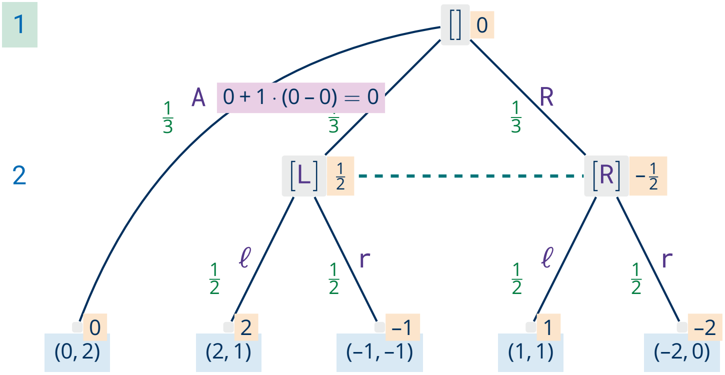
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

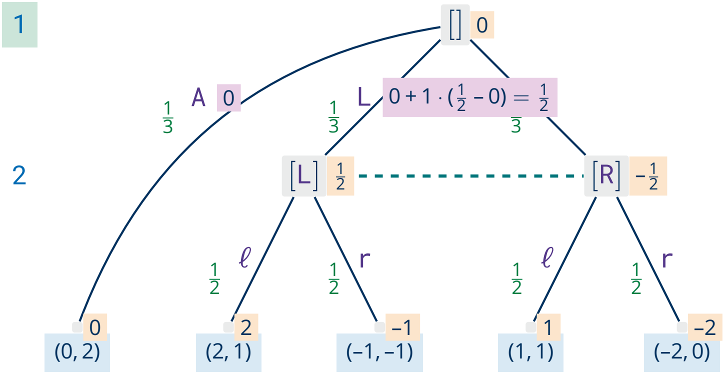
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

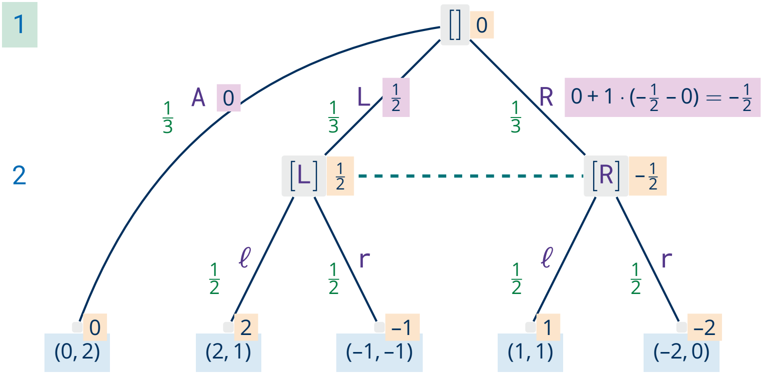
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

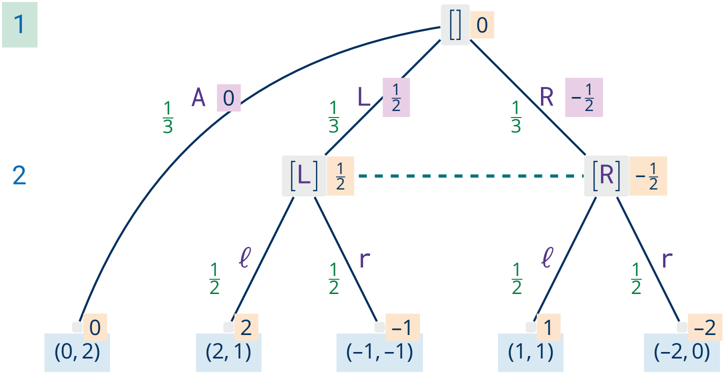
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

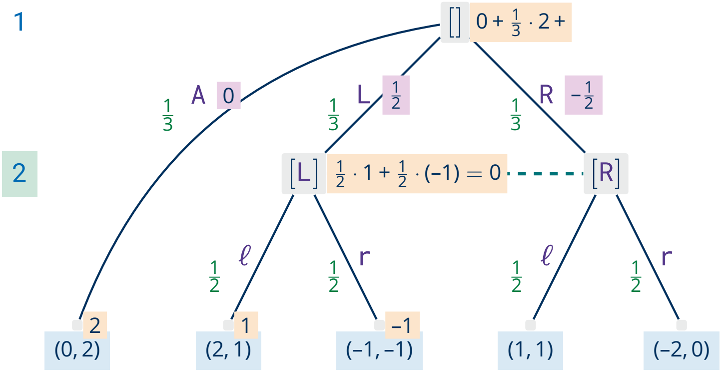
Recall the following extensive-form game G_4:



(2) Traverse game tree for $T = 1, i = 1$

CFR: Example

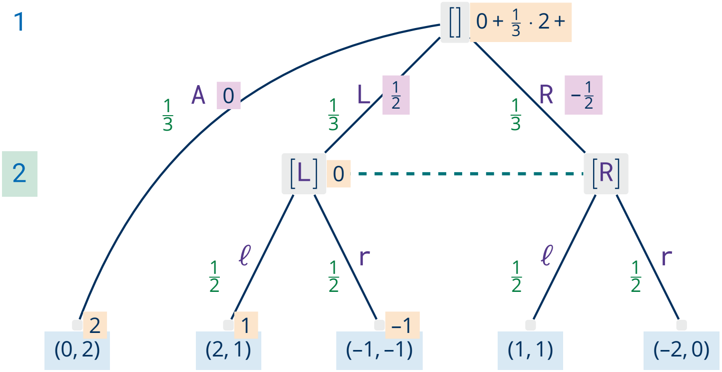
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

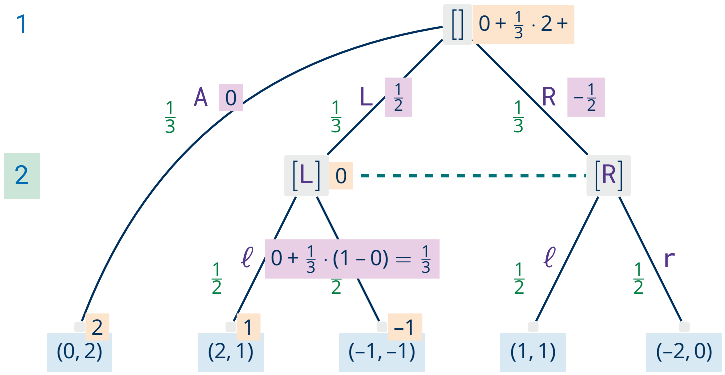
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

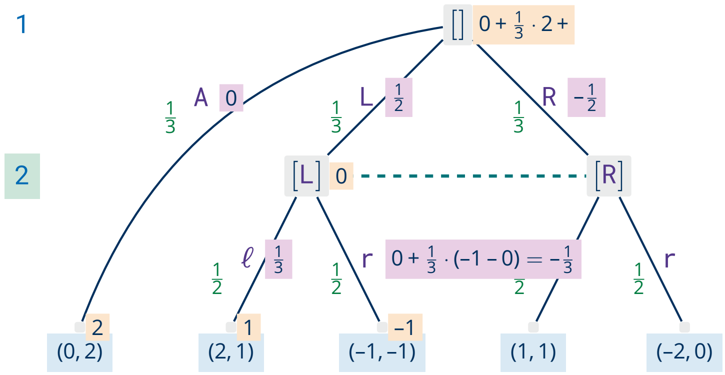
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

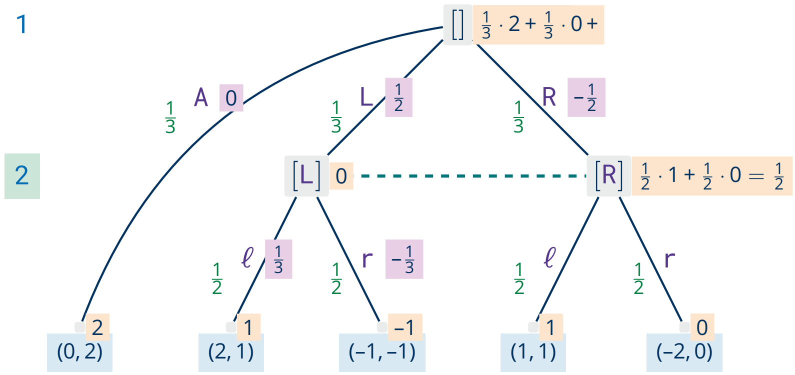
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

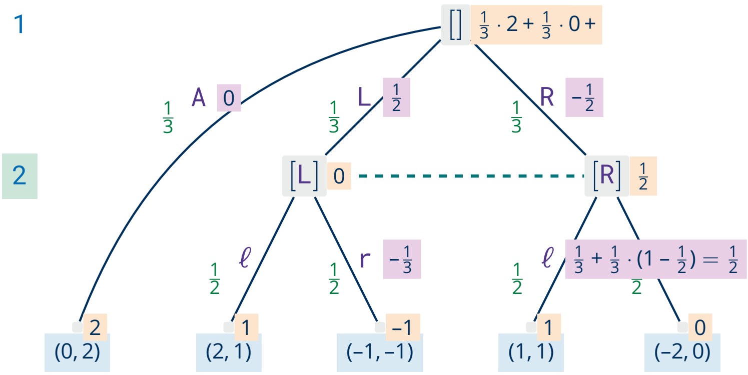
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

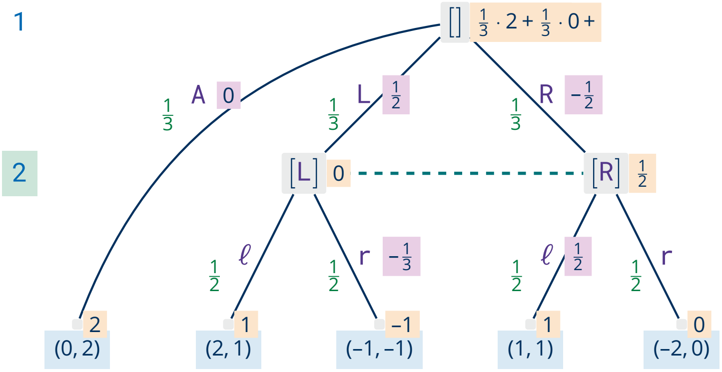
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

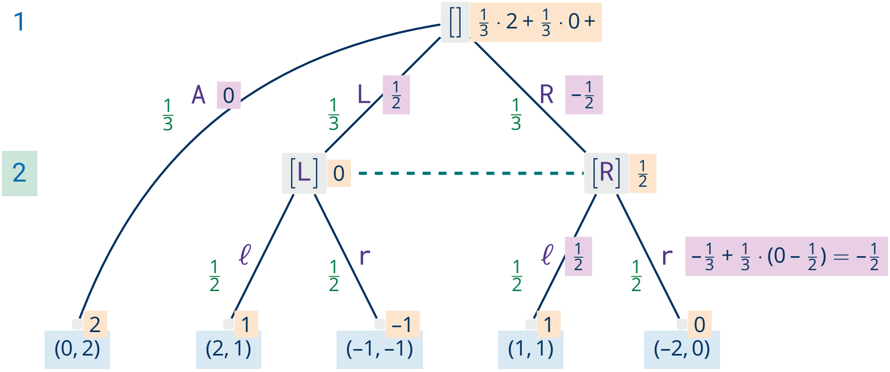
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

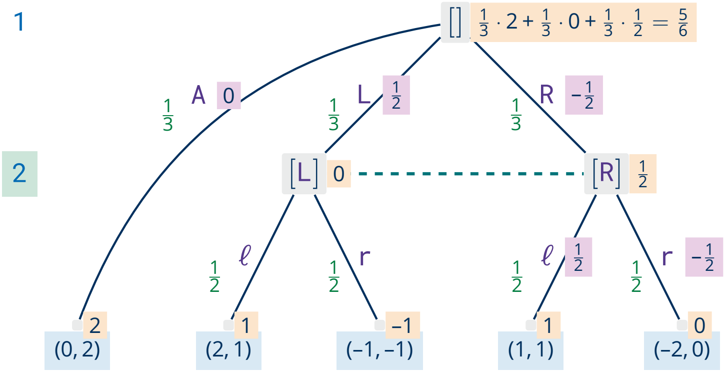
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

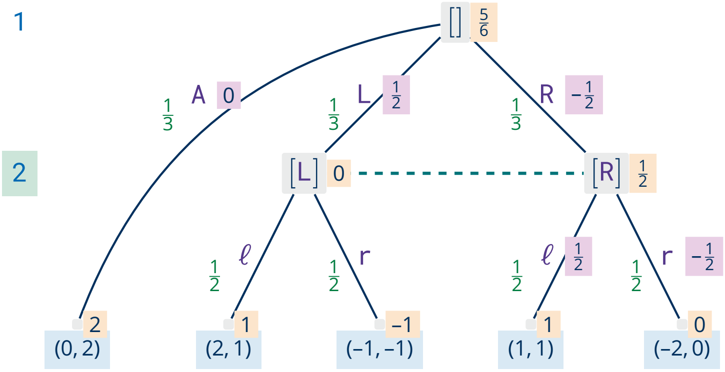
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

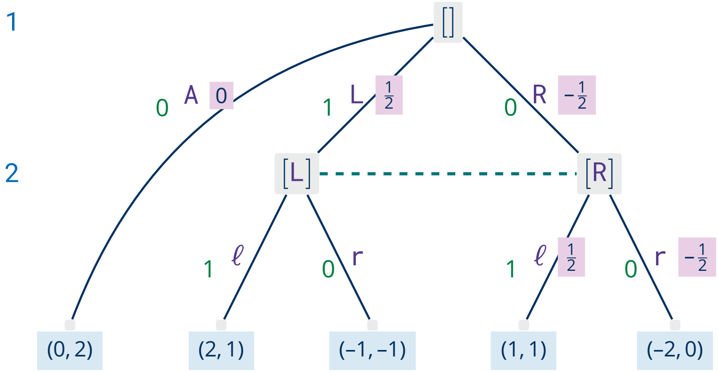
Recall the following extensive-form game G_4:



(3) Traverse game tree for $T = 1, i = 2$

CFR: Example

Recall the following extensive-form game G_4:



(4) Update move probabilities according to regret matching

CFR: Convergence and Correctness

Theorem [Zinkevich, Johanson, Bowling, and Piccione, 2007]

For any extensive-form game with perfect recall, if player i selects actions according to regret matching at information sets, then

$$r_i^T(\mathcal{I}_j) \leq \omega \cdot \sqrt{\left|M_i'\right|} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \left|\left\{\mathcal{I}_j \in \mathcal{I} \,\middle|\, p(\mathcal{I}_j) = i\right\}\right| \cdot \sqrt{\left|M_i'\right|} \cdot \sqrt{T}$$

where $\omega \in \mathbb{R}$ only depends on **u**, and $\left|M_i'\right| := \max_{\mathcal{I}_j \in \mathcal{I}_{p(\mathcal{I}_j)=i}} \left|M_i(\mathcal{I}_j)\right|$.

CFR: Convergence and Correctness

Theorem [Zinkevich, Johanson, Bowling, and Piccione, 2007]

For any extensive-form game with perfect recall, if player i selects actions according to regret matching at information sets, then

$$r_i^T(\mathcal{I}_j) \leq \omega \cdot \sqrt{\left|M_i'\right|} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \left|\left\{\mathcal{I}_j \in \mathcal{I} \mid p(\mathcal{I}_j) = i\right\}\right| \cdot \sqrt{\left|M_i'\right|} \cdot \sqrt{T}$$

where $\omega \in \mathbb{R}$ only depends on $\mathbf{u}$, and $\left|M_i'\right| := \max_{\mathcal{I}_j \in \mathcal{I}_{p(\mathcal{I}_j)=i}} \left|M_i(\mathcal{I}_j)\right|$.

- The bound on overall regret is linear in the number of information sets.

CFR: Convergence and Correctness

Theorem [Zinkevich, Johanson, Bowling, and Piccione, 2007]

For any extensive-form game with perfect recall, if player i selects actions according to regret matching at information sets, then

$$r_i^T(\mathcal{I}_j) \leq \omega \cdot \sqrt{|M_i'|} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \left|\{\mathcal{I}_j \in \mathcal{I} \mid p(\mathcal{I}_j) = i\}\right| \cdot \sqrt{|M_i'|} \cdot \sqrt{T}$$

where $\omega \in \mathbb{R}$ only depends on **u**, and $|M_i'| := \max_{\mathcal{I}_j \in \mathcal{I}_{p(\mathcal{I}_j)=i}} |M_i(\mathcal{I}_j)|$.

- The bound on overall regret is linear in the number of information sets.
- The overall regret grows sublinearly in T, so the **average overall regret** $\bar{R}_i^T := \frac{1}{T} \cdot R_i$ tends to zero as $T \to \infty$.

CFR: Convergence and Correctness

Theorem [Zinkevich, Johanson, Bowling, and Piccione, 2007]

For any extensive-form game with perfect recall, if player i selects actions according to regret matching at information sets, then

$$r_i^T(\mathfrak{I}_j) \leq \omega \cdot \sqrt{|M_i'|} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \left|\{\mathfrak{I}_j \in \mathfrak{I} \mid p(\mathfrak{I}_j) = i\}\right| \cdot \sqrt{|M_i'|} \cdot \sqrt{T}$$

where $\omega \in \mathbb{R}$ only depends on $\mathbf{u}$, and $|M_i'| := \max_{\mathfrak{I}_j \in \mathfrak{I}_{p(\mathfrak{I}_j)=i}} |M_i(\mathfrak{I}_j)|$.

- The bound on overall regret is linear in the number of information sets.
- The overall regret grows sublinearly in T, so the **average overall regret** $\bar{R}_i^T := \frac{1}{T} \cdot R_i$ tends to zero as $T \to \infty$.

Theorem

In any two-player, zero-sum extensive-form game with perfect recall, if both players select actions according to regret matching at information sets, then the average strategy profiles tend to the set of Nash equilibria as $T \to \infty$.

CFR Algorithm: Remarks

- Histories/information sets of `Nature` can be treated in the algorithm via sampling a move from $M_{\texttt{Nature}}(\mathcal{I}_j)$ with the specified distribution.

CFR Algorithm: Remarks

- Histories/information sets of `Nature` can be treated in the algorithm via sampling a move from $M_{\texttt{Nature}}(\mathcal{I}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \ldots, T$ (and for each $i \in P$), the call to **cfr**$([\,], i, t, 1, 1)$ leads to a full traversal of the game tree.

CFR Algorithm: Remarks

- Histories/information sets of `Nature` can be treated in the algorithm via sampling a move from $M_{\texttt{Nature}}(\mathcal{I}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \ldots, T$ (and for each $i \in P$), the call to **cfr**$([\,], i, t, 1, 1)$ leads to a full traversal of the game tree.
- After **solve**(T), the final values of $strategy[j][m]$ can be normalised to obtain the behaviour strategies tending towards Nash equilibria.

CFR Algorithm: Remarks

- Histories/information sets of `Nature` can be treated in the algorithm via sampling a move from $M_{\texttt{Nature}}(\mathcal{I}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \ldots, T$ (and for each $i \in P$), the call to **cfr**$([\,], i, t, 1, 1)$ leads to a full traversal of the game tree.
- After **solve**(T), the final values of $\mathit{strategy}[j][m]$ can be normalised to obtain the behaviour strategies tending towards Nash equilibria.
- Additional techniques, e.g. game abstraction, are used in practice to reduce the number of information sets (per player) to a manageable size.

CFR Algorithm: Remarks

- Histories/information sets of `Nature` can be treated in the algorithm via sampling a move from $M_{\texttt{Nature}}(\mathcal{I}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \ldots, T$ (and for each $i \in P$), the call to **cfr**($[]$, $i, t, 1, 1$) leads to a full traversal of the game tree.
- After **solve**(T), the final values of $strategy[j][m]$ can be normalised to obtain the behaviour strategies tending towards Nash equilibria.
- Additional techniques, e.g. game abstraction, are used in practice to reduce the number of information sets (per player) to a manageable size.
- By using regret matching$^+$ in place of regret matching, we obtain CFR$^+$.

CFR Algorithm: Remarks

- Histories/information sets of `Nature` can be treated in the algorithm via sampling a move from $M_{\texttt{Nature}}(\mathcal{I}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \ldots, T$ (and for each $i \in P$), the call to **cfr**$([\,], i, t, 1, 1)$ leads to a full traversal of the game tree.
- After **solve**(T), the final values of $\mathit{strategy}[j][m]$ can be normalised to obtain the behaviour strategies tending towards Nash equilibria.
- Additional techniques, e.g. game abstraction, are used in practice to reduce the number of information sets (per player) to a manageable size.
- By using regret matching^{+} in place of regret matching, we obtain CFR^{+}.
- CFR^{+} also uses linear weighting to compute average strategies:

$$\bar{\pi}_i^{T,+}(s_j) := \frac{2}{T^2+T} \cdot \sum_{t=1}^{T} \left(t \cdot \pi^t(s_j)\right)$$

CFR Algorithm: Remarks

- Histories/information sets of `Nature` can be treated in the algorithm via sampling a move from $M_{\texttt{Nature}}(\mathcal{I}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \ldots, T$ (and for each $i \in P$), the call to **cfr**$([\,], i, t, 1, 1)$ leads to a full traversal of the game tree.
- After **solve**(T), the final values of $\mathit{strategy}[j][m]$ can be normalised to obtain the behaviour strategies tending towards Nash equilibria.
- Additional techniques, e.g. game abstraction, are used in practice to reduce the number of information sets (per player) to a manageable size.
- By using regret matching$^+$ in place of regret matching, we obtain CFR$^+$.
- CFR$^+$ also uses linear weighting to compute average strategies:

$$\bar{\pi}_i^{T,+}(s_j) := \frac{2}{T^2+T} \cdot \sum_{t=1}^{T} \left(t \cdot \pi^t(s_j)\right)$$

- Bowling et al. [2015] used CFR$^+$ (with additional optimisations) to "essentially weakly solve" heads-up limit hold'em poker.

Conclusion

Summary

- The **regret** is the difference between a player's best possible strategy and their actual strategy.
- A **correlated equilibrium** can be seen as providing players with private signals they can use to best-respond to each other's strategies.
- The **regret matching** algorithm uses self-play to steer play towards the set of correlated equilibria.
- In the case of two-player zero-sum games, regret matching tends towards the set of (mixed) Nash equilibria.
- The **counterfactual regret minimisation** algorithm applies regret matching to every information set of an (imperfect-information) extensive-form game (with perfect recall).

Action Point: Implement CFR$^{(+)}$ and use it to solve Simplified Poker.