

Hannes Strass

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

Counterfactual Regret Minimisation

Lecture 9, 15th Jun 2026 // Algorithmic Game Theory, SS 2026

Previously ...

- 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 is **weakly consistent** whenever the belief system’s probabilities match what is expected from everyone playing according to the behaviour strategy profile.
- An assessment is **consistent** iff it is the limit of a (justifying) sequence of assessments with all-positive-probability strategies.
- An assessment is a **(weak) sequential equilibrium** iff it is both sequentially rational and (weakly) consistent.
- A sequential equilibrium is **perfect** iff every player plays a best response against every opponent profile in the equilibrium’s justifying sequence.

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 $|M_i|^{|P-1(i)|}$ many 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 $|M_i|^{|\rho^{-1}(i)|}$ many pure strategies in the normal-form game.
- Nash equilibria of the normal form do not always respect sequentiality.

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 $|M_i|^{|P-1(i)|}$ many pure strategies in the normal-form game.
- Nash equilibria of the normal form do not always respect sequentiality.

Algorithms for sequential (perfect-information) games?

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

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?

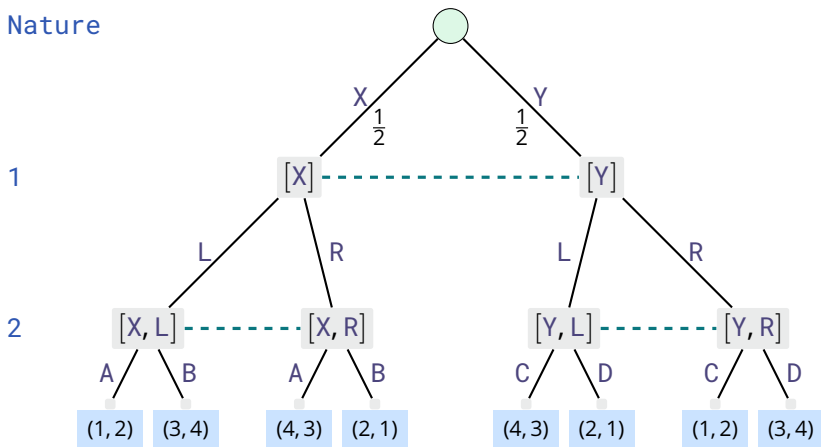
- Incurs an exponential blowup: For every player $i \in P$, there are up to $|M_i|^{|P-1(i)|}$ many pure strategies in the normal-form game.
- Nash equilibria of the normal form do not always respect sequentiality.

Algorithms for sequential (perfect-information) games?

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

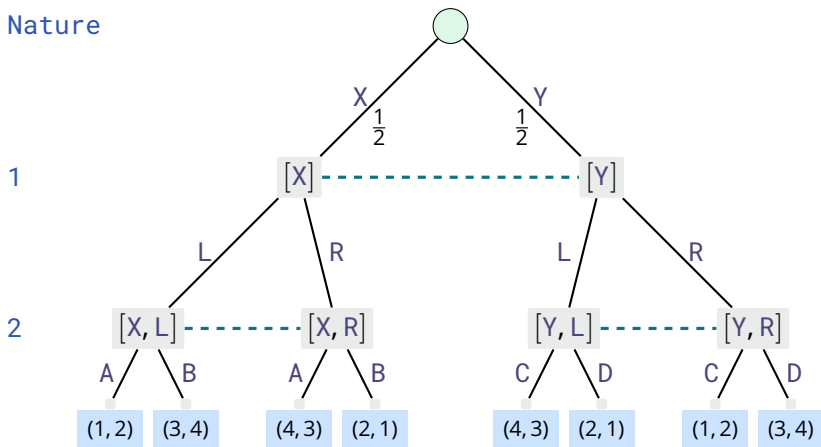
Motivation: Example

Nature



Motivation: Example

Nature



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 $s_j \in S_j$. 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 $s_j \in S_j$. 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 $s_j \in S_j$. 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 $s_j \in S_j$. 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

Regret Matching

Counterfactual Regret Minimisation

Regret Matching

Learning to Play

Learning in Games: General Setting

- A (normal-form) game is played repeatedly for time points $t = 1, 2, \dots$
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^t$ 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 $t = 1, 2, \dots$
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^t$ 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 $t = 1, 2, \dots$
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^t$ 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 $t \rightarrow \infty$ time points, the player “cannot think of” a function $\Phi: \Pi_i \rightarrow \Pi_i$ 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 $t = 1, 2, \dots$
- After the game at time point t has ended, player (say) i has access to all strategy profiles $\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^t$ 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 $t \rightarrow \infty$ time points, the player “cannot think of” a function $\Phi: \Pi_i \rightarrow \Pi_i$ that would strictly increase their payoff in hindsight.

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

Regret Matching: Definition

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

Regret Matching: Definition

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \dots, T$ 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: Definition

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \dots, T$ 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: Definition

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \dots, T$ 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:

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

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

Regret Matching: Definition

In what follows, we assume a fixed normal-form game $G = (P, \mathbf{S}, \mathbf{u})$ to be played at time points $t = 1, 2, \dots, T$ 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:

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

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

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	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	$r_{\text{Cat}}^T(\overline{s_{\text{Cat}}^T})$	$R_{\text{Cat}}^T(\text{Cinema})$	$R_{\text{Cat}}^T(\text{Dancing})$	π_{Cat}^{T+1}
-----	---	---	-----------------------------------	------------------------------------	--------------------------

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	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	$r_{\text{Cat}}^T(\overline{s_{\text{Cat}}^T})$	$R_{\text{Cat}}^T(\text{Cinema})$	$R_{\text{Cat}}^T(\text{Dancing})$	π_{Cat}^{T+1}
1	(Cinema, Dancing)	5	0	5	{Cinema \mapsto 0, Dancing \mapsto 1}

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	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	$r_{\text{Cat}}^T(\overline{s_{\text{Cat}}^T})$	$R_{\text{Cat}}^T(\text{Cinema})$	$R_{\text{Cat}}^T(\text{Dancing})$	π_{Cat}^{T+1}
1	(Cinema, Dancing)	5	0	5	{Cinema \mapsto 0, Dancing \mapsto 1}
2	(Dancing, Cinema)	10	10	5	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }

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	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	$r_{\text{Cat}}^T(\overline{s_{\text{Cat}}^T})$	$R_{\text{Cat}}^T(\text{Cinema})$	$R_{\text{Cat}}^T(\text{Dancing})$	π_{Cat}^{T+1}
1	(Cinema, Dancing)	5	0	5	{Cinema \mapsto 0, Dancing \mapsto 1}
2	(Dancing, Cinema)	10	10	5	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }
3	(Cinema, Dancing)	5	10	10	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }

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	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	$r_{\text{Cat}}^T(\overline{s_{\text{Cat}}^T})$	$R_{\text{Cat}}^T(\text{Cinema})$	$R_{\text{Cat}}^T(\text{Dancing})$	π_{Cat}^{T+1}
1	(Cinema, Dancing)	5	0	5	{Cinema \mapsto 0, Dancing \mapsto 1}
2	(Dancing, Cinema)	10	10	5	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }
3	(Cinema, Dancing)	5	10	10	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }
4	(Cinema, Cinema)	-10	10	0	{Cinema \mapsto 1, Dancing \mapsto 0}

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	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	$r_{\text{Cat}}^T(\overline{s_{\text{Cat}}^T})$	$R_{\text{Cat}}^T(\text{Cinema})$	$R_{\text{Cat}}^T(\text{Dancing})$	π_{Cat}^{T+1}
1	(Cinema, Dancing)	5	0	5	{Cinema \mapsto 0, Dancing \mapsto 1}
2	(Dancing, Cinema)	10	10	5	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }
3	(Cinema, Dancing)	5	10	10	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }
4	(Cinema, Cinema)	-10	10	0	{Cinema \mapsto 1, Dancing \mapsto 0}
5	(Cinema, Dancing)	5	10	5	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }

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	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	$r_{\text{Cat}}^T(s_{\text{Cat}}^T)$	$R_{\text{Cat}}^T(\text{Cinema})$	$R_{\text{Cat}}^T(\text{Dancing})$	π_{Cat}^{T+1}
1	(Cinema, Dancing)	5	0	5	{Cinema \mapsto 0, Dancing \mapsto 1}
2	(Dancing, Cinema)	10	10	5	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }
3	(Cinema, Dancing)	5	10	10	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }
4	(Cinema, Cinema)	-10	10	0	{Cinema \mapsto 1, Dancing \mapsto 0}
5	(Cinema, Dancing)	5	10	5	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }
6	(Cinema, Cinema)	-10	10	-5	{Cinema \mapsto 1, 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

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

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 $T \rightarrow \infty$.

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

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

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 $T \rightarrow \infty$.

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

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

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 $T \rightarrow \infty$.

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

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

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 $T \rightarrow \infty$.

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*.

↪ 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 $R_i^T = (R_i^T(\text{Cinema}), R_i^T(\text{Dancing}))$ for $i \in \{\text{Cat}, \text{Dee}\}$.

T	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	R_{Cat}^T	R_{Dee}^T	π_{Cat}^{T+1}	π_{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 $R_i^T = (R_i^T(\text{Cinema}), R_i^T(\text{Dancing}))$ for $i \in \{\text{Cat}, \text{Dee}\}$.

T	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	R_{Cat}^T	R_{Dee}^T	π_{Cat}^{T+1}	π_{Dee}^{T+1}
1	(Cinema, Dancing)	(0, 5)	(5, 0)	{Cinema \mapsto 0, Dancing \mapsto 1}	{Cinema \mapsto 1, 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 $R_i^T = (R_i^T(\text{Cinema}), R_i^T(\text{Dancing}))$ for $i \in \{\text{Cat}, \text{Dee}\}$.

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

Regret Matching in Self-Play: Example

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

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

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

Regret Matching in Self-Play: Example

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

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

T	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	R_{Cat}^T	R_{Dee}^T	π_{Cat}^{T+1}	π_{Dee}^{T+1}
1	(Cinema, Dancing)	(0, 5)	(5, 0)	{Cinema \mapsto 0, Dancing \mapsto 1}	{Cinema \mapsto 1, Dancing \mapsto 0}
2	(Dancing, Cinema)	(10, 5)	(5, 10)	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }	{Cinema \mapsto $\frac{1}{3}$, Dancing \mapsto $\frac{2}{3}$ }
3	(Cinema, Dancing)	(10, 10)	(10, 10)	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }
4	(Dancing, Dancing)	(5, 10)	(0, 10)	{Cinema \mapsto $\frac{1}{3}$, Dancing \mapsto $\frac{2}{3}$ }	{Cinema \mapsto 0, 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 $R_i^T = (R_i^T(\text{Cinema}), R_i^T(\text{Dancing}))$ for $i \in \{\text{Cat}, \text{Dee}\}$.

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

Regret Matching in Self-Play: Example

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

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

T	$\mathbf{s}^T = (s_{\text{Cat}}^T, s_{\text{Dee}}^T)$	R_{Cat}^T	R_{Dee}^T	π_{Cat}^{T+1}	π_{Dee}^{T+1}
1	(Cinema, Dancing)	(0, 5)	(5, 0)	{Cinema \mapsto 0, Dancing \mapsto 1}	{Cinema \mapsto 1, Dancing \mapsto 0}
2	(Dancing, Cinema)	(10, 5)	(5, 10)	{Cinema \mapsto $\frac{2}{3}$, Dancing \mapsto $\frac{1}{3}$ }	{Cinema \mapsto $\frac{1}{3}$, Dancing \mapsto $\frac{2}{3}$ }
3	(Cinema, Dancing)	(10, 10)	(10, 10)	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }	{Cinema \mapsto $\frac{1}{2}$, Dancing \mapsto $\frac{1}{2}$ }
4	(Dancing, Dancing)	(5, 10)	(0, 10)	{Cinema \mapsto $\frac{1}{3}$, Dancing \mapsto $\frac{2}{3}$ }	{Cinema \mapsto 0, Dancing \mapsto 1}
5	(Cinema, Dancing)	(5, 15)	(5, 10)	{Cinema \mapsto $\frac{1}{4}$, Dancing \mapsto $\frac{3}{4}$ }	{Cinema \mapsto $\frac{1}{3}$, Dancing \mapsto $\frac{2}{3}$ }
6	(Dancing, Dancing)	(0, 15)	(-5, 10)	{Cinema \mapsto 0, Dancing \mapsto 1}	{Cinema \mapsto 0, 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 $R_i^T = (R_i^T(\text{Cinema}), R_i^T(\text{Dancing}))$ for $i \in \{\text{Cat}, \text{Dee}\}$.

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

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

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

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

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

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

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

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

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 $T \rightarrow \infty$ iff $\bar{\boldsymbol{\varphi}}^T$ 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

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

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

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

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 $T \rightarrow \infty$, the pair $(\bar{\pi}_1^T, \bar{\pi}_2^T)$ 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:

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

Regret Matching⁺

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

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

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

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

Regret Matching⁺

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

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

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

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

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

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

Regret Matching⁺

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

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

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

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

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

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

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

$P(h' | h, \sigma)$ is the probability that h' is reached when playing σ from h on:

- $P(h | h, \sigma) = 1$ for all $h \in H$,
- $P(\square | h, \sigma) = 0$ for all $h \neq \square$, and
- $P([h'; m] | h, \sigma) = \sigma_{p(\mathcal{J}_{h'})}(m | \mathcal{J}_{h'}) \cdot P(h' | h, \sigma)$.

Remember, Remember

Recall

$P(h' | h, \sigma)$ is the probability that h' is reached when playing σ from h on:

- $P(h | h, \sigma) = 1$ for all $h \in H$,
- $P(\square | h, \sigma) = 0$ for all $h \neq \square$, and
- $P([h'; m] | h, \sigma) = \sigma_{p(\mathcal{J}_{h'})}(m | \mathcal{J}_{h'}) \cdot P(h' | h, \sigma)$.

Recall

The probability of reaching information set \mathcal{J}_j when playing σ is thus

$$P(\mathcal{J}_j | \sigma) := \sum_{h \in \mathcal{J}_j} P(h | \sigma) \quad \text{where } P(h | \sigma) \text{ denotes } P(h | \square, \sigma)$$

Remember, Remember

Recall

$P(h' | h, \sigma)$ is the probability that h' is reached when playing σ from h on:

- $P(h | h, \sigma) = 1$ for all $h \in H$,
- $P(\square | h, \sigma) = 0$ for all $h \neq \square$, and
- $P([h'; m] | h, \sigma) = \sigma_{p(\mathcal{J}_{h'})}(m | \mathcal{J}_{h'}) \cdot P(h' | h, \sigma)$.

Recall

The probability of reaching information set \mathcal{J}_j when playing σ is thus

$$P(\mathcal{J}_j | \sigma) := \sum_{h \in \mathcal{J}_j} P(h | \sigma) \quad \text{where } P(h | \sigma) \text{ denotes } P(h | \square, \sigma)$$

Recall

Player i 's expected utility of playing σ when history h has been reached is

$$U_i(\sigma | h) := \sum_{z \in Z} P(z | h, \sigma) \cdot u_i(z)$$

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

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

$$P(\mathcal{J}_j | \sigma_{-i}) := \sum_{h \in \mathcal{J}_j} P(h | \sigma_{-i})$$

We **counterfactually** assume that i **intentionally** played to reach \mathcal{J}_j .

Towards Counterfactual Regret

Definition

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

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

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

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

$$P(\mathcal{J}_j | \sigma_{-i}) := \sum_{h \in \mathcal{J}_j} P(h | \sigma_{-i})$$

3. The **counterfactual utility** of **playing to reach** \mathcal{J}_j and then playing σ is

$$U_i(\sigma | \mathcal{J}_j) = \frac{\sum_{h \in \mathcal{J}_j} P(h | \sigma_{-i}) \cdot U_i(\sigma | h)}{P(\mathcal{J}_j | \sigma_{-i})} = \frac{\sum_{h \in \mathcal{J}_j} P(h | \sigma_{-i}) \cdot \sum_{z \in Z} P(z | h, \sigma) \cdot u_i(z)}{P(\mathcal{J}_j | \sigma_{-i})}$$

We **counterfactually** assume that i **intentionally** played to reach \mathcal{J}_j .

Counterfactual Regret

Definition

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

1. Recall: The possible moves of i in \mathcal{J}_j are

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

Counterfactual Regret

Definition

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

1. Recall: The possible moves of i in \mathcal{J}_j are

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

2. For behaviour strategy profile σ and move $m \in M_i(\mathcal{J}_j)$, define modified profile $\langle \sigma \rangle_m^{\mathcal{J}_j}$ to be just like σ , except that in \mathcal{J}_j it always chooses m .

Counterfactual Regret

Definition

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

1. Recall: The possible moves of i in \mathcal{J}_j are

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

2. For behaviour strategy profile σ and move $m \in M_i(\mathcal{J}_j)$, define modified profile $\langle \sigma \rangle_m^{\mathcal{J}_j}$ to be just like σ , except that in \mathcal{J}_j it always chooses m .

3. The **immediate counterfactual regret** at time T is then defined by

$$r_i^T(\mathcal{J}_j) := \max_{m^* \in M_i(\mathcal{J}_j)} \sum_{t=1}^T P(\mathcal{J}_j \mid \sigma_{-i}^t) \cdot \left(U_i \left(\langle \sigma^t \rangle_{m^*}^{\mathcal{J}_j} \mid \mathcal{J}_j \right) - U_i(\sigma^t \mid \mathcal{J}_j) \right)$$

for any sequence $(\sigma^t)_{t=1}^T$ of behaviour strategy profiles.

Counterfactual Regret

Definition

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

1. Recall: The possible moves of i in \mathcal{J}_j are

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

2. For behaviour strategy profile σ and move $m \in M_i(\mathcal{J}_j)$, define modified profile $\langle \sigma \rangle_m^{\mathcal{J}_j}$ to be just like σ , except that in \mathcal{J}_j it always chooses m .

3. The **immediate counterfactual regret** at time T is then defined by

$$r_i^T(\mathcal{J}_j) := \max_{m^* \in M_i(\mathcal{J}_j)} \sum_{t=1}^T P(\mathcal{J}_j \mid \sigma_{-i}^t) \cdot \left(U_i \left(\langle \sigma^t \rangle_{m^*}^{\mathcal{J}_j} \mid \mathcal{J}_j \right) - U_i(\sigma^t \mid \mathcal{J}_j) \right)$$

for any sequence $(\sigma^t)_{t=1}^T$ of behaviour strategy profiles.

Key Feature: $r_i^T(\mathcal{J}_j)$ can be minimised by controlling only $\sigma_i(\mathcal{J}_j): M_i(\mathcal{J}_j) \rightarrow [0, 1]$.

Overall Regret \leq Immediate Regret

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

$$R_i^T = \max_{\sigma_i^* \in \Sigma_i} \sum_{t=1}^T (U_i(\sigma_i^*, \sigma_{-i}^t) - U_i(\sigma^t))$$

where $U_i(\sigma)$ denotes $U_i(\sigma \mid \cdot)$.

Overall Regret \leq Immediate Regret

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

$$R_i^T = \max_{\sigma_i^* \in \Sigma_i} \sum_{t=1}^T (U_i(\sigma_i^*, \sigma_{-i}^t) - U_i(\sigma^t))$$

where $U_i(\sigma)$ denotes $U_i(\sigma \mid \cdot)$.

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

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

$$R_i^T \leq \sum_{J_j \in p^{-1}(i)} [r_j^T(J_j)]^+$$

Overall Regret \leq Immediate Regret

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

$$R_i^T = \max_{\sigma_i^* \in \Sigma_i} \sum_{t=1}^T (U_i(\sigma_i^*, \sigma_{-i}^t) - U_i(\sigma^t))$$

where $U_i(\sigma)$ denotes $U_i(\sigma \mid \cdot)$.

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

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

$$R_i^T \leq \sum_{J_j \in p^{-1}(i)} [r_j^T(J_j)]^+$$

Thus: Minimising immediate regret in each J_j minimises overall regret.

Regret Matching at Information Sets

Definition

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

1. Let $\mathcal{J}_j \in p^{-1}(i)$ and $m \in M_i(\mathcal{J}_j)$. The **accumulated regret** of move m is

$$R_i^T(\mathcal{J}_j, m) := \sum_{t=1}^T P(\mathcal{J}_j | \sigma_{-i}^t) \cdot \left(U_i \left(\langle \sigma^t \rangle_m^{\mathcal{J}_j} \mid \mathcal{J}_j \right) - U_i(\sigma^t \mid \mathcal{J}_j) \right)$$

Regret Matching at Information Sets

Definition

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

1. Let $\mathcal{J}_j \in p^{-1}(i)$ and $m \in M_i(\mathcal{J}_j)$. The **accumulated regret** of move m is

$$R_i^T(\mathcal{J}_j, m) := \sum_{t=1}^T P(\mathcal{J}_j | \sigma_{-i}^t) \cdot \left(U_i(\langle \sigma^t \rangle_m^{\mathcal{J}_j} | \mathcal{J}_j) - U_i(\sigma^t | \mathcal{J}_j) \right)$$

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

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

CFR: Algorithm (1)

Initialisation of global variables:

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

CFR: Algorithm (1)

Initialisation of global variables:

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

Main Loop:

```
function solve( $T$ ) {  
  foreach  $t \in \{1, 2, \dots, T\}$  do {  
    foreach  $i \in \{1, 2\}$  do {  
      cfr( $\square, i, t, 1, 1$ )  
    }  
  }  
}
```

CFR: Algorithm (2)

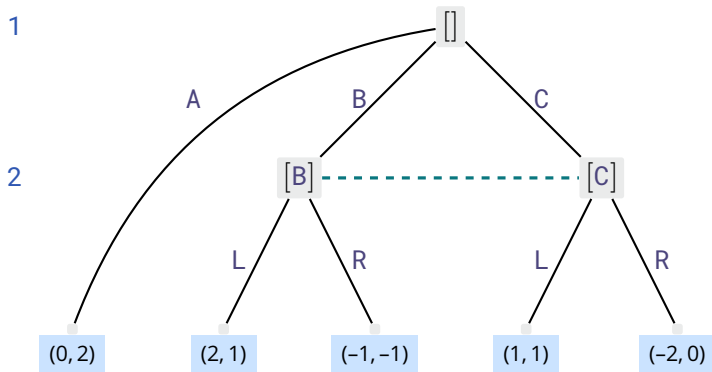
```

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  $h \in \mathcal{J}_j$ 
    foreach  $m \in M_{p(\mathcal{J}_j)}(\mathcal{J}_j)$  do {  $v'_h[j][m] := 0$  } // initialise payoffs of single moves
    foreach  $m \in M_{p(\mathcal{J}_j)}(\mathcal{J}_j)$  do {
        if TURN( $h$ ) = 1 then {  $v'_h[j][m] := \text{cfr}([h; m], i, t, \text{profile}[t][j][m] \cdot p_1, p_2)$  }
        else {  $v'_h[j][m] := \text{cfr}([h; m], i, t, p_1, \text{profile}[t][j][m] \cdot p_2)$  }
         $v_h := v_h + \text{profile}[t][j][m] \cdot v'_h[j][m]$  // accumulate currently expected payoff
    }
    if TURN( $h$ ) =  $i$  then { // players minimise immediate regret of own moves
         $r^+ := 0$  // initialise sum of positive regrets
        for  $m \in M_i(\mathcal{J}_j)$  do { // update values needed for regret matching
             $\text{regret}[j][m] := \text{regret}[j][m] + p_{3-i} \cdot (v'_h[m] - v_h)$  // update accumulated cf regret
             $\text{strategy}[j][m] := \text{strategy}[j][m] + p_i \cdot \text{profile}[t][j][m]$  // update "frequency" of move
             $r^+ := r^+ + [\text{regret}[j][m]]^+$  // accumulate positive regret sum for normalisation
        }
        if  $r^+ > 0$  then { foreach  $m \in M_i(\mathcal{J}_j)$  do { // apply regret matching at  $\mathcal{J}_j$ 
             $\text{profile}[t+1][j][m] := [\text{regret}[j][m]]^+ / r^+$  } }
        else { foreach  $m \in M_i(\mathcal{J}_j)$  do {
             $\text{profile}[t+1][j][m] := 1 / |M_i(\mathcal{J}_j)|$  } } }
    }
    return  $v_h$  }

```

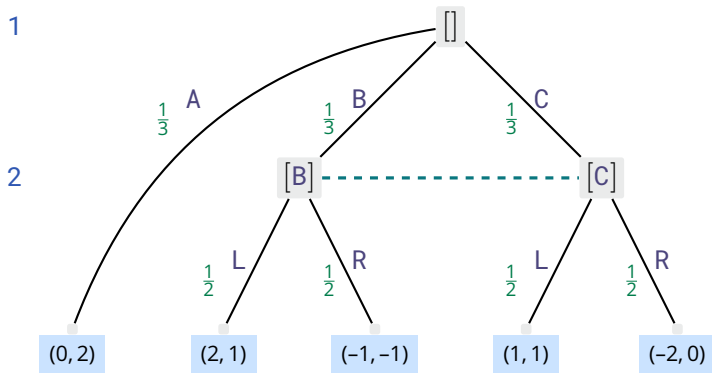
CFR: Example

Recall the following extensive-form game G_4 :



CFR: Example

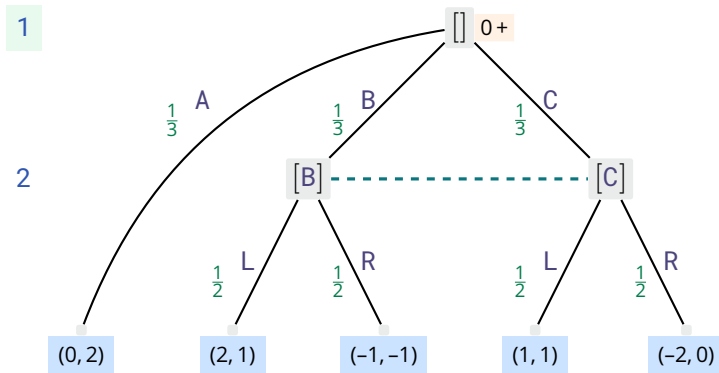
Recall the following extensive-form game G_4 :



(1) Initialise move probabilities by uniform distributions

CFR: Example

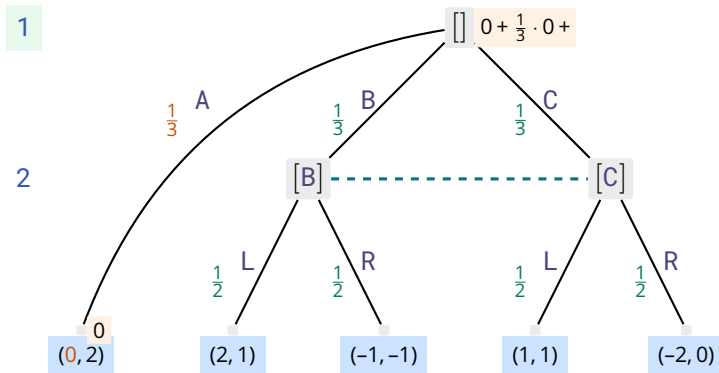
Recall the following extensive-form game G_4 :



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

CFR: Example

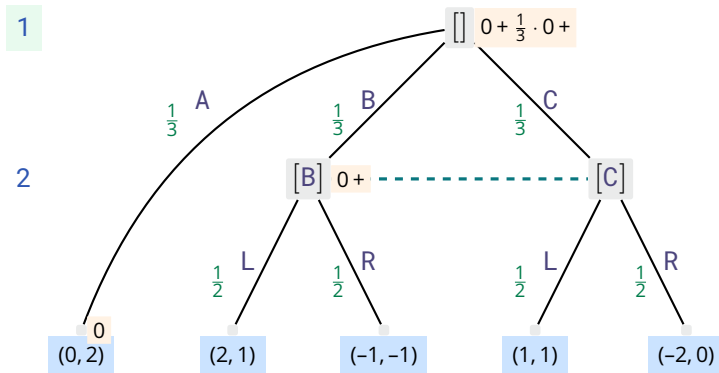
Recall the following extensive-form game G_4 :



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

CFR: Example

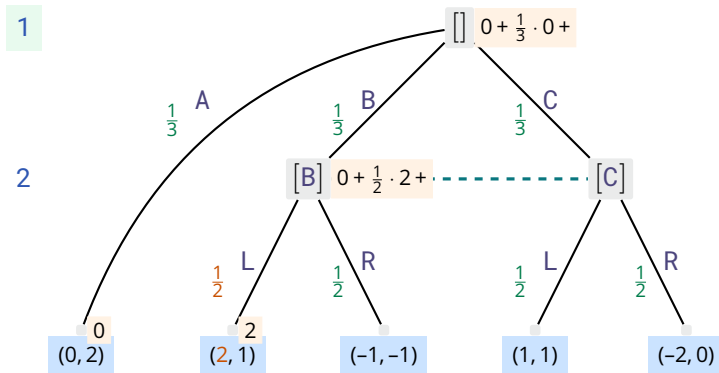
Recall the following extensive-form game G_4 :



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

CFR: Example

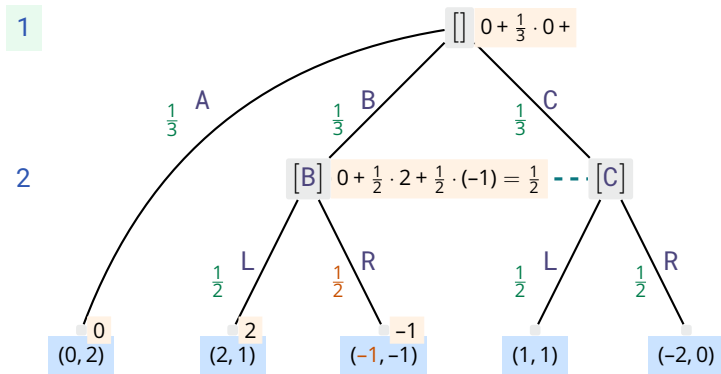
Recall the following extensive-form game G_4 :



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

CFR: Example

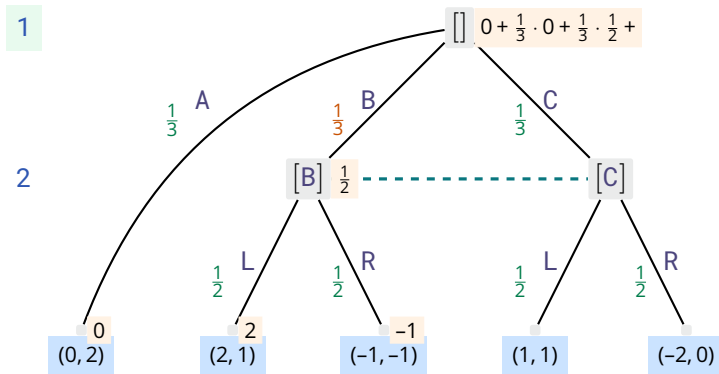
Recall the following extensive-form game G_4 :



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

CFR: Example

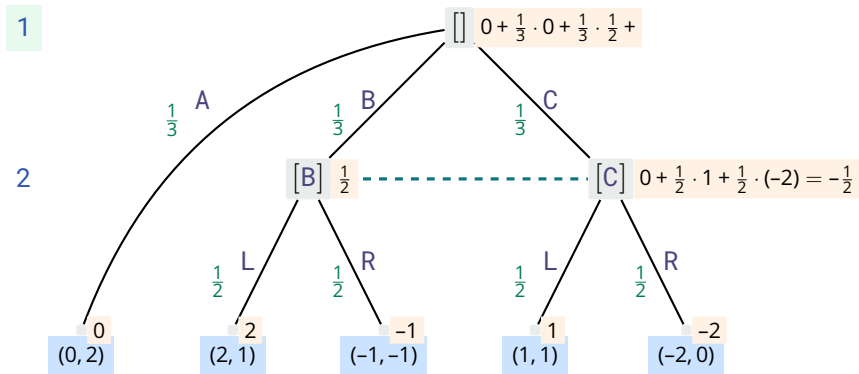
Recall the following extensive-form game G_4 :



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

CFR: Example

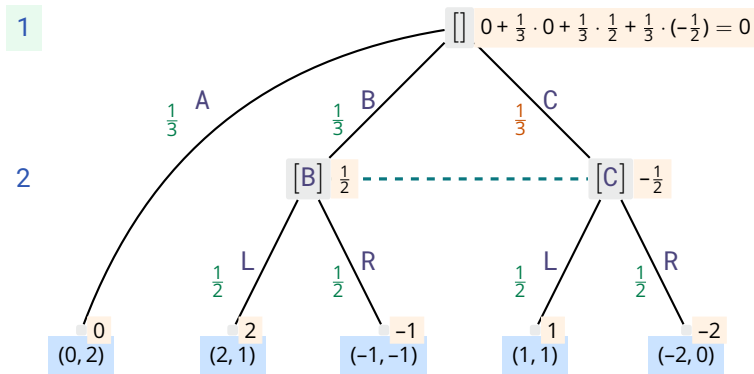
Recall the following extensive-form game G_4 :



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

CFR: Example

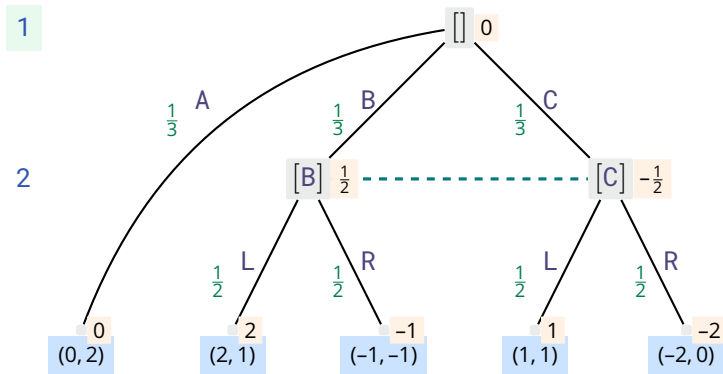
Recall the following extensive-form game G_4 :



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

CFR: Example

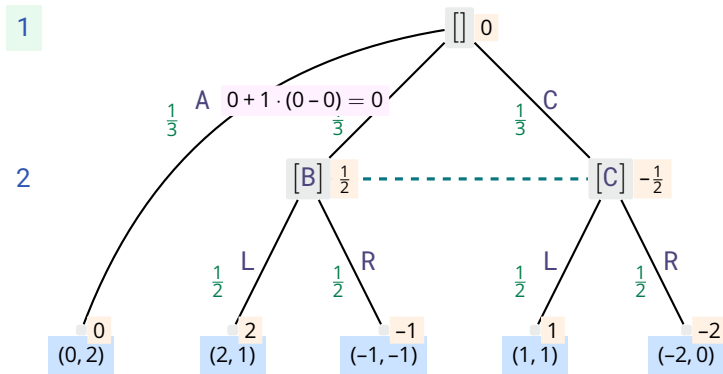
Recall the following extensive-form game G_4 :



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

CFR: Example

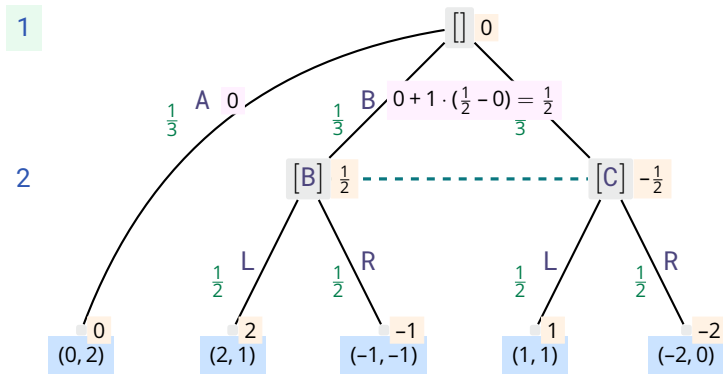
Recall the following extensive-form game G_4 :



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

CFR: Example

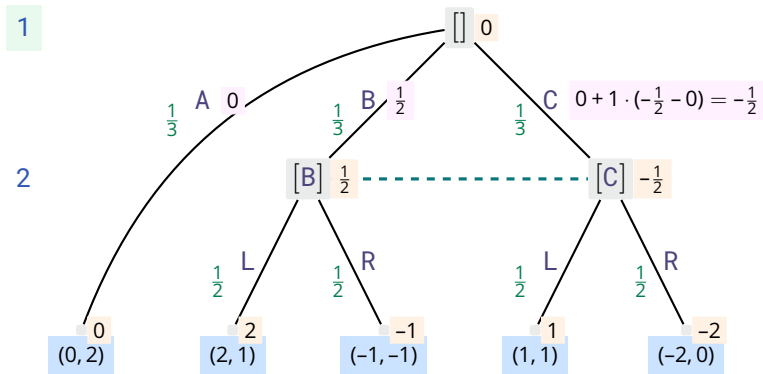
Recall the following extensive-form game G_4 :



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

CFR: Example

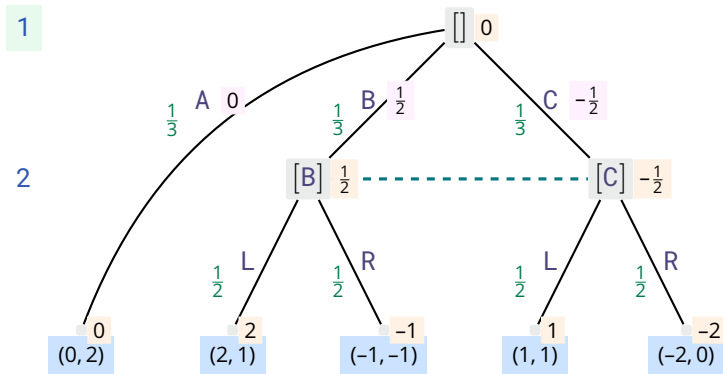
Recall the following extensive-form game G_4 :



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

CFR: Example

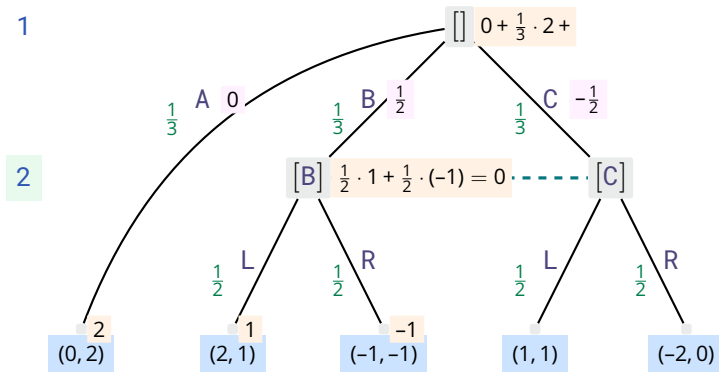
Recall the following extensive-form game G_4 :



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

CFR: Example

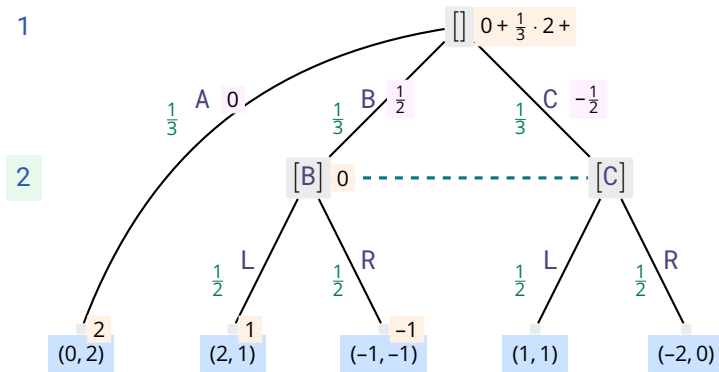
Recall the following extensive-form game G_4 :



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

CFR: Example

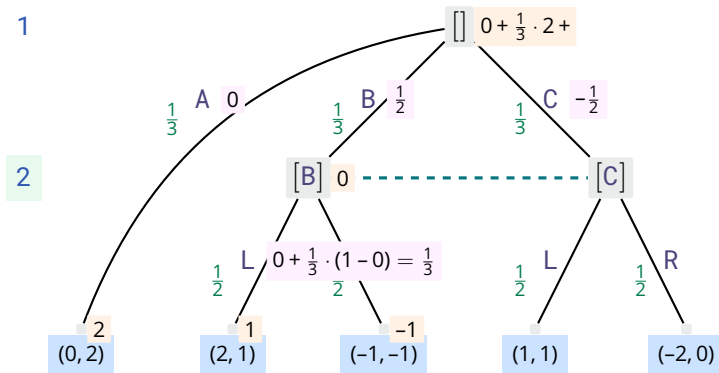
Recall the following extensive-form game G_4 :



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

CFR: Example

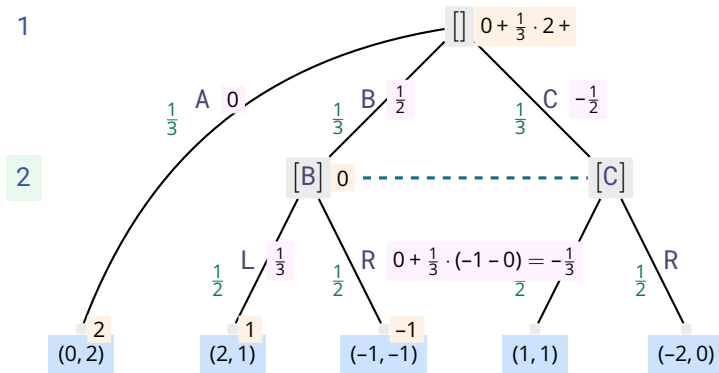
Recall the following extensive-form game G_4 :



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

CFR: Example

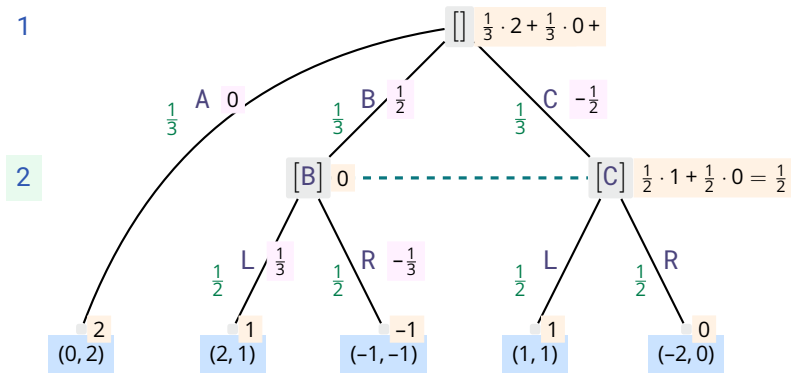
Recall the following extensive-form game G_4 :



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

CFR: Example

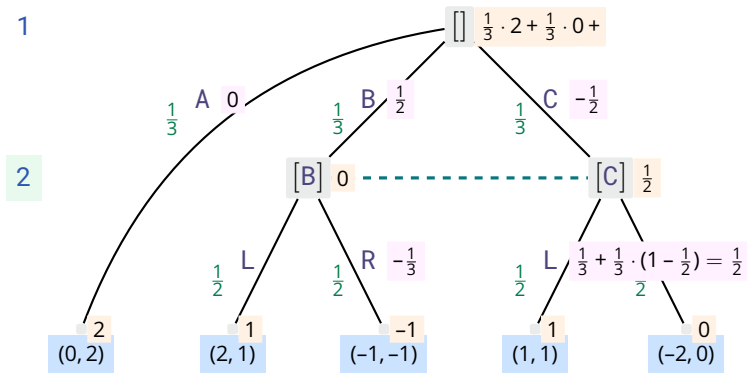
Recall the following extensive-form game G_4 :



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

CFR: Example

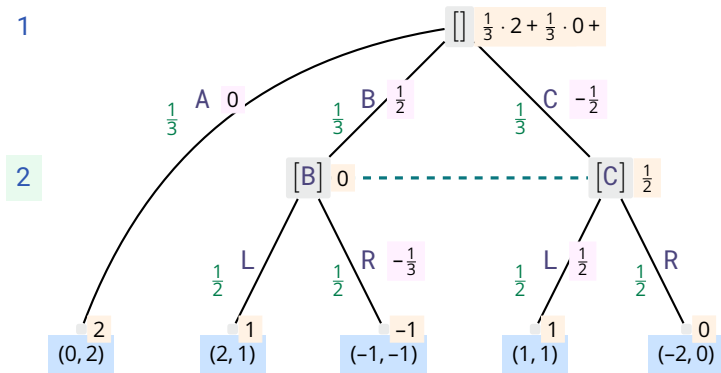
Recall the following extensive-form game G_4 :



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

CFR: Example

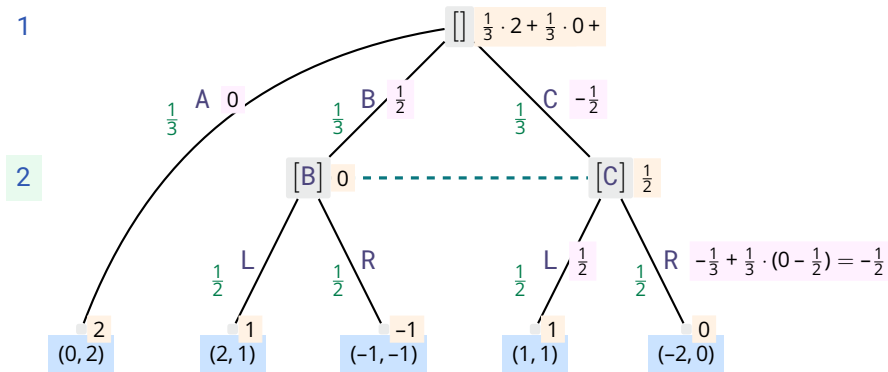
Recall the following extensive-form game G_4 :



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

CFR: Example

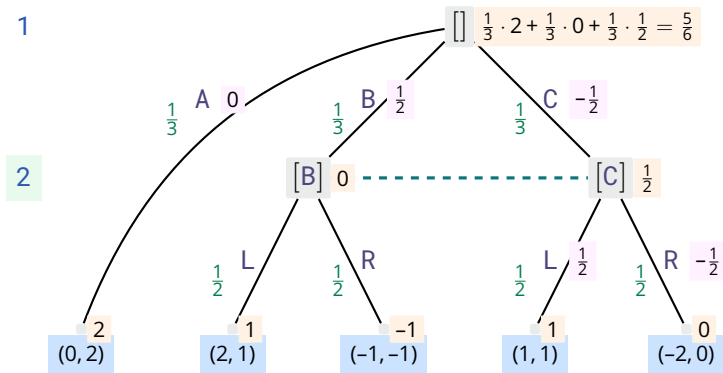
Recall the following extensive-form game G_4 :



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

CFR: Example

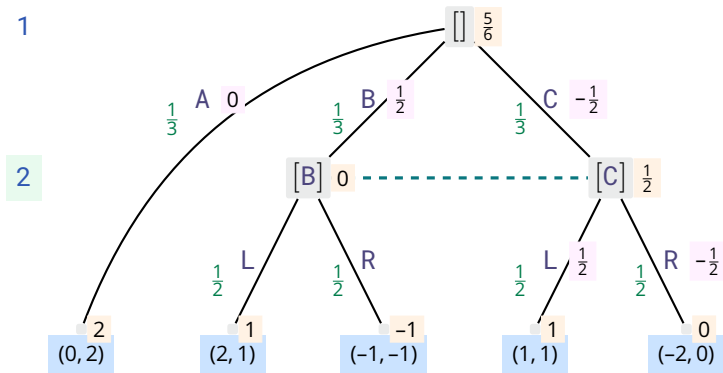
Recall the following extensive-form game G_4 :



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

CFR: Example

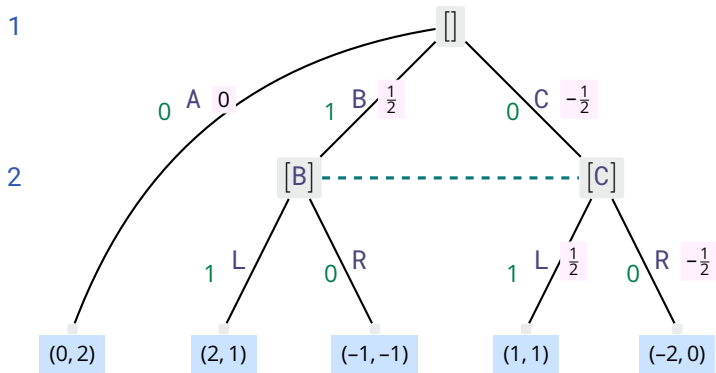
Recall the following extensive-form game G_4 :



(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

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

$$r_i^T(\mathcal{J}_j) \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \cdot |p^{-1}(i)|$$

where $\omega \in \mathbb{R}$ only depends on \mathbf{u} , and $\mu_i := \max_{\mathcal{J}_j \in p^{-1}(i)} |M_i(\mathcal{J}_j)|$.

CFR: Convergence

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

$$r_i^T(\mathcal{J}_j) \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \cdot |p^{-1}(i)|$$

where $\omega \in \mathbb{R}$ only depends on \mathbf{u} , and $\mu_i := \max_{\mathcal{J}_j \in p^{-1}(i)} |M_i(\mathcal{J}_j)|$.

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

CFR: Convergence

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

$$r_i^T(\mathcal{J}_j) \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \cdot |p^{-1}(i)|$$

where $\omega \in \mathbb{R}$ only depends on \mathbf{u} , and $\mu_i := \max_{\mathcal{J}_j \in p^{-1}(i)} |M_i(\mathcal{J}_j)|$.

- The bound on overall regret is **linear** in the number of information sets.
- The overall regret grows **sublinearly** in T , thus we obtain:

CFR: Convergence

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

$$r_i^T(\mathcal{J}_j) \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \quad \text{whence} \quad R_i^T \leq \omega \cdot \sqrt{\mu_i} \cdot \sqrt{T} \cdot |p^{-1}(i)|$$

where $\omega \in \mathbb{R}$ only depends on \mathbf{u} , and $\mu_i := \max_{\mathcal{J}_j \in p^{-1}(i)} |M_i(\mathcal{J}_j)|$.

- The bound on overall regret is **linear** in the number of information sets.
- The overall regret grows **sublinearly** in T , thus we obtain:

Corollary

For the **average overall regret** $\bar{R}_i^T := \frac{1}{T} \cdot R_i$, we have $\lim_{T \rightarrow \infty} \bar{R}_i^T = 0$.

CFR: Correctness

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 $T \rightarrow \infty$.

CFR: Correctness

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 $T \rightarrow \infty$.

Why Nash equilibria and not sequential equilibria?

- In zero-sum games, the minimax value of the game is unique (even when there are different equilibria).

CFR: Correctness

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 $T \rightarrow \infty$.

Why Nash equilibria and not sequential equilibria?

- In zero-sum games, the minimax value of the game is unique (even when there are different equilibria).
- Off-equilibrium behaviour can be adjusted without changing the outcome.

CFR: Correctness

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 $T \rightarrow \infty$.

Why Nash equilibria and not sequential equilibria?

- In zero-sum games, the minimax value of the game is unique (even when there are different equilibria).
- Off-equilibrium behaviour can be adjusted without changing the outcome.
- In any case, regret matching **at every information set** already strives for sequential rationality ...

CFR: Correctness

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 $T \rightarrow \infty$.

Why Nash equilibria and not sequential equilibria?

- In zero-sum games, the minimax value of the game is unique (even when there are different equilibria).
- Off-equilibrium behaviour can be adjusted without changing the outcome.
- In any case, regret matching **at every information set** already strives for sequential rationality ...
- ...but there are no known results on convergence to sequential equilibria.

CFR: Correctness

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 $T \rightarrow \infty$.

Why Nash equilibria and not sequential equilibria?

- In zero-sum games, the minimax value of the game is unique (even when there are different equilibria).
- Off-equilibrium behaviour can be adjusted without changing the outcome.
- In any case, regret matching **at every information set** already strives for sequential rationality ...
- ...but there are no known results on convergence to sequential equilibria.
- For general-sum games, convergence is only guaranteed for more permissive equilibrium concepts (e.g. correlated equilibria for extensive-form games).

CFR Algorithm: Remarks

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

CFR Algorithm: Remarks

- Histories/information sets of **Nature** can be treated in the algorithm via sampling a move from $M_{\text{Nature}}(\mathcal{J}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \dots, T$ (and for each $i \in P$), the call to **cfr**($\square, i, t, 1, 1$) 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 $M_{\text{Nature}}(\mathcal{J}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \dots, T$ (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 $M_{\text{Nature}}(\mathcal{J}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \dots, T$ (and for each $i \in P$), the call to **cfr**($\square, i, t, 1, 1$) 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.

CFR Algorithm: Remarks

- Histories/information sets of **Nature** can be treated in the algorithm via sampling a move from $M_{\text{Nature}}(\mathcal{I}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \dots, T$ (and for each $i \in P$), the call to **cfr**($\square, i, t, 1, 1$) 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 $M_{\text{Nature}}(\mathcal{J}_j)$ with the specified distribution.
- At each time step $t = 1, 2, \dots, T$ (and for each $i \in P$), the call to **cfr**($\square, i, t, 1, 1$) 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⁺ also uses linear weighting to compute average strategies:

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

CFR Algorithm: Remarks

- Histories/information sets of **Nature** can be treated in the algorithm via sampling a move from $M_{\text{Nature}}(j)$ with the specified distribution.
- At each time step $t = 1, 2, \dots, T$ (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⁺ also uses linear weighting to compute average strategies:

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

- 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.
- The **regret matching** algorithm uses self-play to steer average play towards the set of correlated equilibria.
- In the case of two-player zero-sum games, regret matching of both players tends towards the set of (mixed) Nash equilibria.
- The **counterfactual regret minimisation** algorithm applies regret matching to every information set of an extensive-form game with perfect recall.
- For two-player zero-sum games, the counterfactual regret minimisation (CFR) algorithm's computed strategies tend to the set of Nash equilibria.

Action Point

Implement CFR⁽⁺⁾ and use it to solve Simplified Poker.