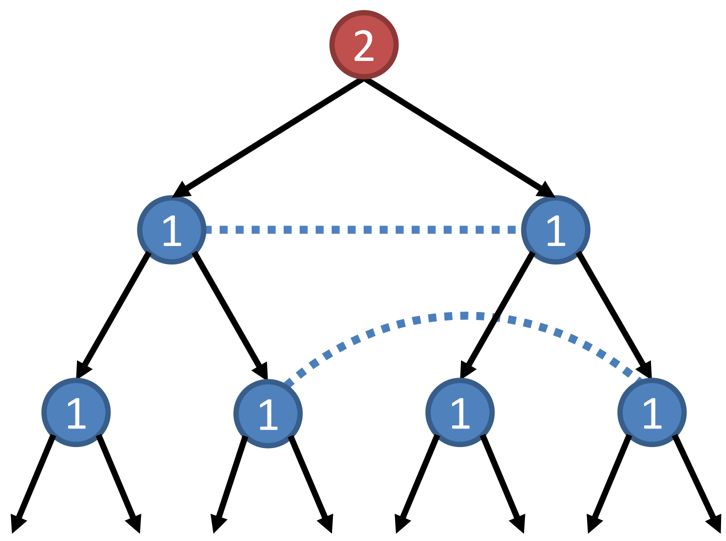
Extensive-Form Games

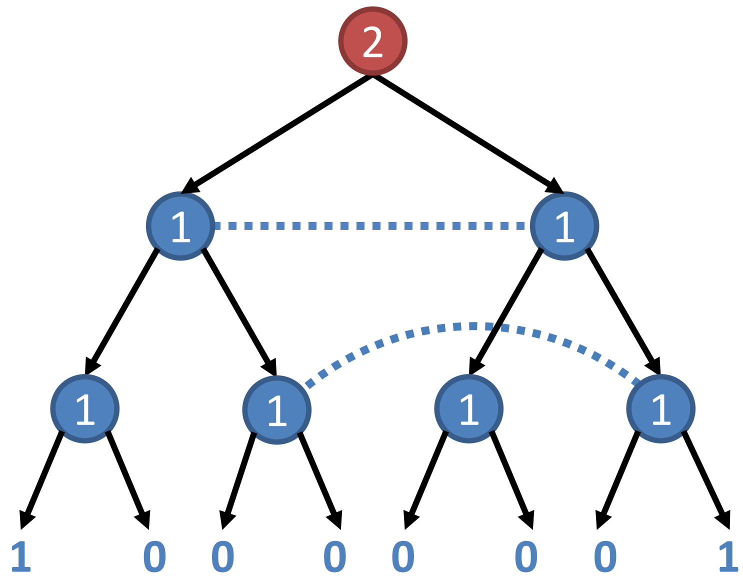
Brian Zhang

Extensive-Form Games

- Game represented by a **tree**
- Can capture sequential and simultaneous moves
- Private information
- We assume **perfect recall**: no player forgets what the player knew earlier



Perfect Recall and Sequences



P1 doesn't know the action of **P2**

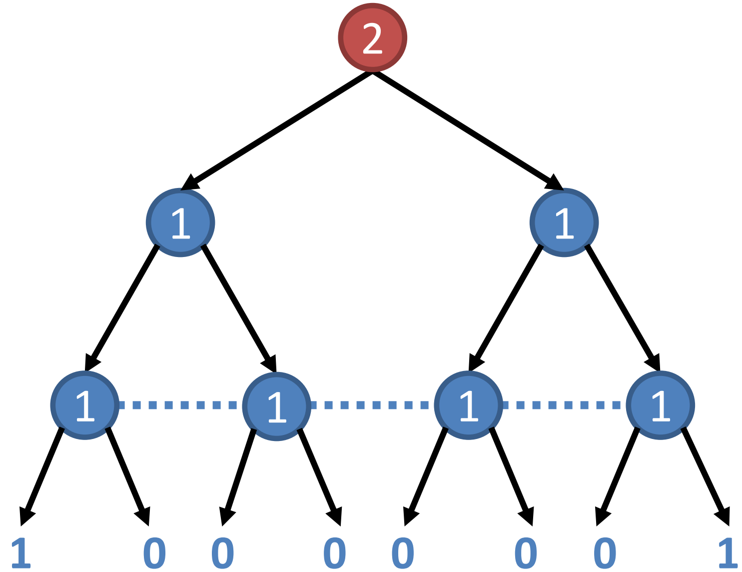
P1 now learns the action of **P2** if **P1** played **left**

✓ Perfect recall!

Defn: At a history h, the *sequence* $\sigma_i(h)$ (or *player history*) of player i is the ordered list of previous infosets encountered by that player and actions played at those infosets (conventionally, not including the infoset at h itself)

Player i has **perfect recall** if $\sigma_i(h) = \sigma_i(h')$ whenever h, h' share an infoset

Perfect Recall and Sequences



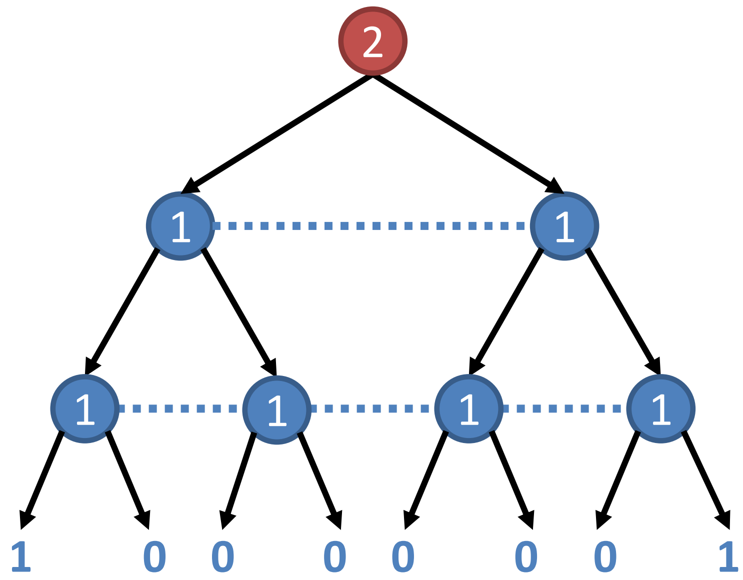
P1 knows the action of **P2**

✗ **P1** has forgotten the action of **P2**

Defn: At a history h, the *sequence* $\sigma_i(h)$ (or *player history*) of player i is the ordered list of previous infosets encountered by that player and actions played at those infosets (conventionally, not including the infoset at h itself)

Player i has **perfect recall** if $\sigma_i(h) = \sigma_i(h')$ whenever h, h' share an infoset

Perfect Recall and Sequences



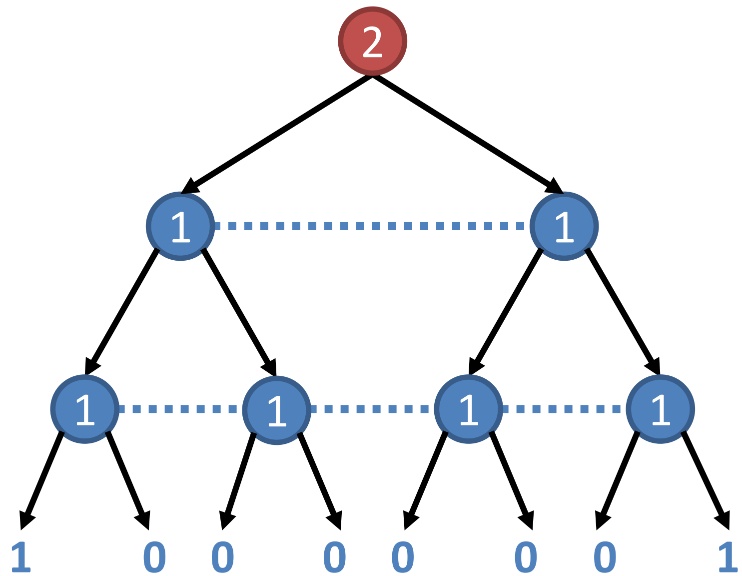
P1 does not know the action of **P2**

❌ **P1** has forgotten **P1**'s own action

Defn: At a history h, the *sequence* $\sigma_i(h)$ (or *player history*) of player i is the ordered list of previous infosets encountered by that player and actions played at those infosets (conventionally, not including the infoset at h itself)

Player i has **perfect recall** if $\sigma_i(h) = \sigma_i(h')$ whenever h, h' share an infoset

Strategies and Kuhn's Theorem



P1 does not know the action of **P2**

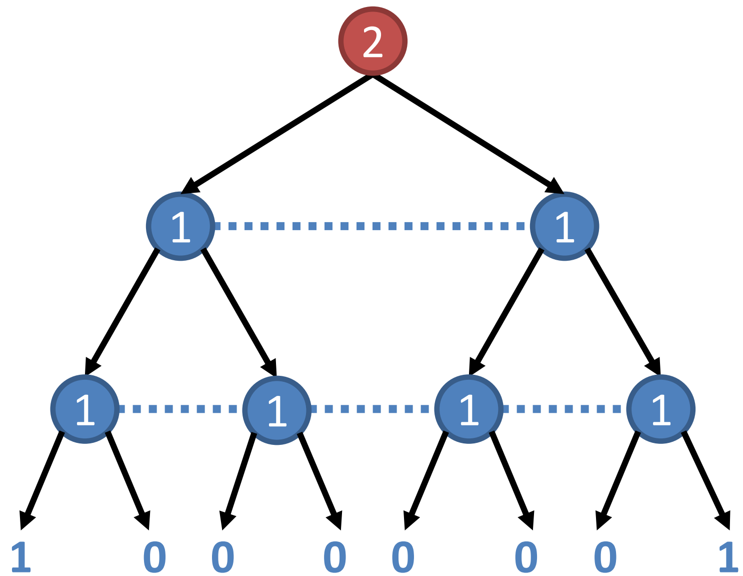
❌ **P1** has forgotten **P1**'s own action

Pure strategy $\pi_i \in \Pi_i$ of player i = map $\pi_i : \mathcal{J}_i \to A$
Mixed strategy = distribution $\mu_i \in \Delta(\Pi_\mathrm{i})$
Behavioral strategy = a mixed strategy that mixes **independently** at every infoset
map $\pi_i : \mathcal{J}_i \to \Delta(A)$

Q: Are behavioral strategies as expressive as mixed strategies?
A: ("Kuhn's theorem", *Contrib. Theory of Games* 1950) Yes, but *only for players w/ perfect recall*

for all $\mu_i \in \Delta(\Pi_i)$, exists $\pi_i : \mathcal{J}_i \to \Delta(A)$ s.t. $u_i(\pi_i, \pi_{-i}) = u_i(\mu_i, \pi_{-i}) \; \forall \pi_{-i} \in \times_{j \neq i} \Pi_j$

Strategies and Kuhn's Theorem



Mixed strategy

$$\mu_1 = \frac{1}{2}(L,L) + \frac{1}{2}(R,R)$$

has $u_1(\mu_1, \pi_2) = \frac{1}{2} \ \ \forall \pi_2$

but *no behavioral strategy* has this

Pure strategy $\pi_i \in \Pi_i$ of player i = map $\pi_i : \mathcal{J}_i \to A$
Mixed strategy = distribution $\mu_i \in \Delta(\Pi_\mathrm{i})$
Behavioral strategy = a mixed strategy that mixes **independently** at every infoset
map $\pi_i : \mathcal{J}_i \to \Delta(A)$

Q: Are behavioral strategies as expressive as mixed strategies?
A: ("Kuhn's theorem", *Contrib. Theory of Games* 1950) Yes, but *only for players w/ perfect recall*

for all $\mu_i \in \Delta(\Pi_i)$, exists $\pi_i : \mathcal{J}_i \to \Delta(A)$ s.t. $u_i(\pi_i, \pi_{-i}) = u_i(\mu_i, \pi_{-i}) \ \forall \pi_{-i} \in \times_{j \neq i} \Pi_j$

Digression: Why trees?

A: Because even computing the optimal strategy in a one-player DAG-form games with perfect recall ("finite-horizon POMDPs") is PSPACE-hard [Papadimitriou & Tsitsiklis, *Math of OR* 1987]

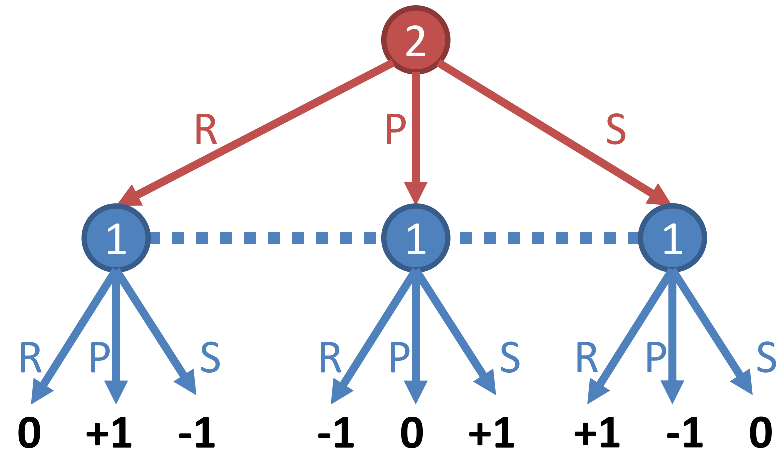
(#histories can be exponentially larger than #states; optimal strategy can be history dependent)

Extensive-form games are

- expressive enough to capture real-world settings (usually sequential & imperfect-information), and yet
- "simple" enough to allow **positive** results

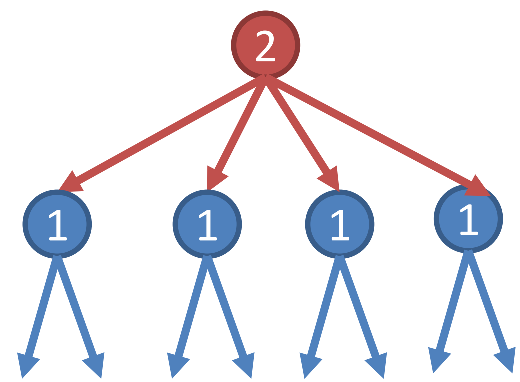
Normal ⟷ Extensive Form

	R	P	S
R	0	-1	1
P	1	0	-1
S	-1	1	0



LLLL				
LLLR				
LLRL				
⋮				
RRRR				

Combinatorial explosion!



Tree-Form Decision Making

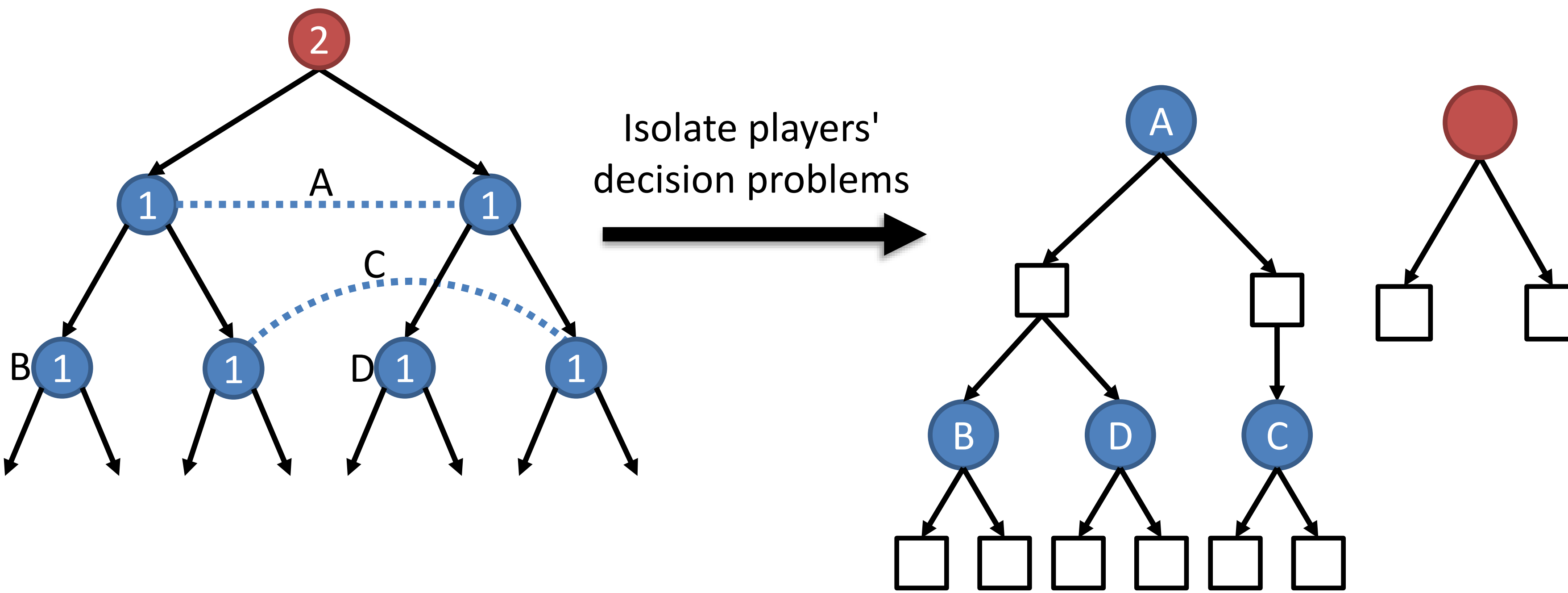
Game tree

Each node belongs to a specific player (or chance, not pictured)

Tree-form (Sequential) decision problems
aka. sequence-form decision problems
aka. treeplexes

Represents the game from viewpoint of one player

This is the representation for this lecture



Tree-Form Decision Making

First attempt:

Assign local probabilities at each decision point

✓ Set of strategies is convex

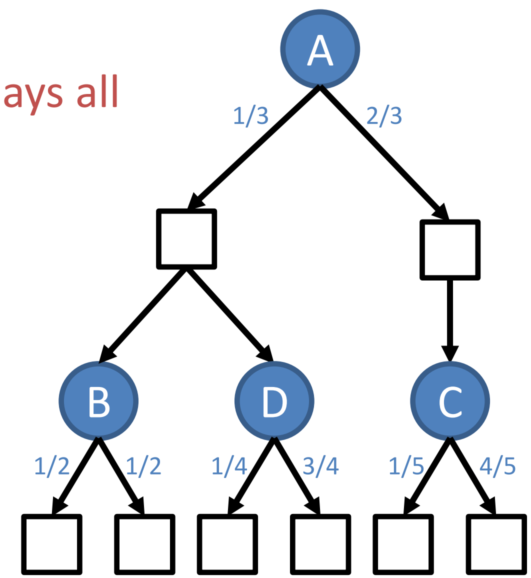
✗ Expected utility of game is **not** bilinear

$$\sum_{z \in Z} u_1(z)\, c(z) \prod_{i \in [n]} \left(\prod_{\substack{ha \preccurlyeq z \\ h \text{ belongs to } i}} \pi_i(a|h) \right)$$

set of terminal nodes (tree leaves)

probability that *chance (nature)* plays all actions on the path to z

probability that Player i plays all actions on the path to z



Sequence Form: A Way to Generate a Polynomial-Sized LP in the Size of the Tree

Given $\pi_i : \mathcal{J}_i \to \Delta(A)$, the **sequence form** $\boldsymbol{x}_i \in \mathbb{R}^{\Sigma_i}$ of π_i is

$$x_i[\sigma] = \prod_{(I,a)\in\sigma} \pi_i(a|I)$$

set of sequences

$x_1[A\ell] = {}^1/_3, \qquad x_1[Ar] = {}^2/_3$

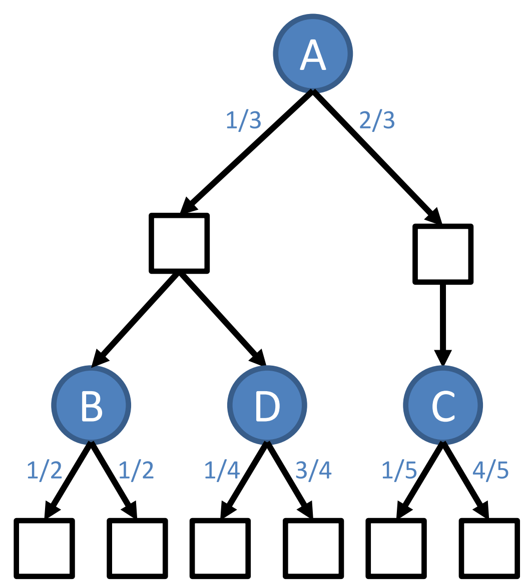
$x_1[B\ell] = x_1[Br] = {}^1/_3 \cdot {}^1/_2 = {}^1/_6$

$x_1[D\ell] = {}^1/_3 \cdot {}^1/_4 = {}^1/_{12}$

$x_1[Dr] = {}^1/_3 \cdot {}^3/_4 = {}^1/_4$

$x_1[C\ell] = {}^2/_3 \cdot {}^1/_5 = {}^2/_{15}$

$x_1[Cr] = {}^2/_3 \cdot {}^4/_5 = {}^8/_{15}$



Sequence Form: A Way to Generate a Polynomial-Sized LP in the Size of the Tree

The set of sequence-form strategies is a **convex polytope**!

$$X_i \coloneqq \left\{ \boldsymbol{x}_i \in \mathbb{R}_{\geq 0}^{\Sigma_i} \;:\; x_i[\emptyset] = 1, \quad \sum_{a \in A} x_i[Ia] = x_i[\sigma_i(I)] \;\; \forall I \in \mathcal{J} \right\}$$

$$= \left\{ \boldsymbol{x}_i \in \mathbb{R}_{\geq 0}^{\Sigma_i} \;:\; \boldsymbol{F}_i \boldsymbol{x}_i = \boldsymbol{f}_i \right\}$$

$\sigma_i(I)$: common sequence of all histories in I

The utility of player i is **linear** in i's sequence-form strategy!

$$u_i(x) = \sum_{z \in Z} u_i(z) \cdot c(z) \cdot \prod_{i \in [n]} x_i[\sigma_i(z)]$$

Z: set of terminal nodes (tree leaves)

$c(z)$: probability that *chance (nature)* plays all actions on the path to z

Two-player zero-sum case:

$$u_i(x, y) = \sum_{z \in Z} u_i(z) \cdot c(z) \cdot x[\sigma_1(z)] \cdot y[\sigma_2(z)] = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$$

Sequence Form LP

$$\max_{\boldsymbol{x} \in X} \ \min_{\boldsymbol{y} \in Y} \ \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$$

Two-player zero-sum case:

$$u_i(x, y) = \sum_{z \in Z} u_i(z) \cdot c(z) \cdot x[\sigma_1(z)] \cdot y[\sigma_2(z)] = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$$

Sequence Form LP

$$\max_{\boldsymbol{x}} \left\{ \begin{array}{l} \min_{\boldsymbol{y}} \ \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y} \\ \text{s.t. } \boldsymbol{F}_2\boldsymbol{y} = \boldsymbol{f}_2, \\ \qquad \boldsymbol{y} \geq \boldsymbol{0} \end{array} \right.$$

$$\text{s.t. } \boldsymbol{F}_1\boldsymbol{x} = \boldsymbol{f}_1, \quad \boldsymbol{x} \geq \boldsymbol{0}$$

LP duality

$$\max_{\boldsymbol{v}} \ \boldsymbol{f}_2^\top \boldsymbol{v}$$
$$\text{s.t. } \boldsymbol{A}^\top \boldsymbol{x} \geq \boldsymbol{F}_2^\top \boldsymbol{v}$$

Two-player zero-sum case:

$$u_i(x, y) = \sum_{z \in Z} u_i(z) \cdot c(z) \cdot x[\sigma_1(z)] \cdot y[\sigma_2(z)] = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{y}$$

Sequence Form LP

$$\max_{\boldsymbol{x},\boldsymbol{v}} \boldsymbol{f}_2^\top \boldsymbol{v}$$

$$\text{s.t. } \boldsymbol{F}_1\boldsymbol{x} = \boldsymbol{f}_1, \quad \boldsymbol{x} \geq \boldsymbol{0}, \quad \boldsymbol{A}^\top\boldsymbol{x} \geq \boldsymbol{F}_2^\top \boldsymbol{v}$$

Two-player zero-sum extensive-form games can be solved in polynomial time!

$$\max_{\boldsymbol{v}} \ \boldsymbol{f}_2^\top \boldsymbol{v}$$

$$\text{s.t. } \boldsymbol{A}^\top\boldsymbol{x} \geq \boldsymbol{F}_2^\top \boldsymbol{v}$$

Two-player zero-sum case:

$$u_i(x, y) = \sum_{z \in Z} u_i(z) \cdot c(z) \cdot x[\sigma_1(z)] \cdot y[\sigma_2(z)] = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{y}$$

Regret Minimization on Sequence-Form Strategy Sets

Recall: Regret Minimization

for $t = 1, \ldots, T$:

- Agent chooses a *sequence-form strategy* $\boldsymbol{x}^t \in X \subset \mathbb{R}^n$
- Environment chooses a *utility vector* $\boldsymbol{u}^t \in [0,1]^n$
- Agent observes $\boldsymbol{u}^t$ and gets utility $\langle \boldsymbol{u}^t, \boldsymbol{x}^t \rangle$

Agent goal: Minimize *regret*.
"How well do we do against best, fixed strategy in hindsight?"

$$R^T := \max_{\hat{\boldsymbol{x}} \in X} \left\{ \sum_{t=1}^{T} \langle \boldsymbol{u}^t, \hat{\boldsymbol{x}} \rangle \right\} - \sum_{t=1}^{T} \langle \boldsymbol{u}^t, \boldsymbol{x}^t \rangle$$

Maximum utility that was achievable by the **best fixed** strategy in hindsight

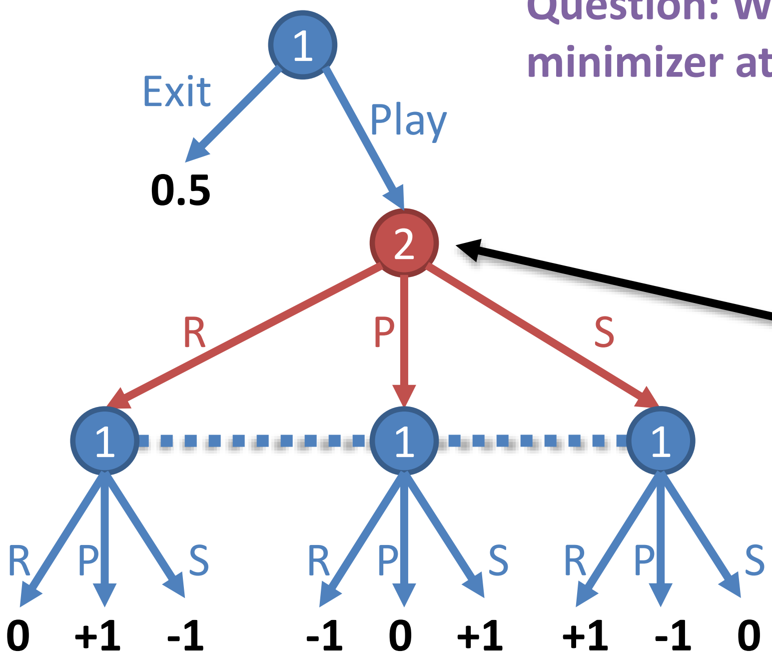
Utility that was actually accumulated

Goal: have R^T grow sublinearly with respect to time T, e.g., $R^T = \text{poly}(n) \cdot \sqrt{T}$

If we can do this, we can learn equilibria!

Counterfactual Regret Minimization (CFR): The Gist

✸ **IDEA:** Run one regret minimizer at each information set!



Question: What utility should we give to each infoset's regret minimizer at each time step?

~~**Attempt 1:** Use "Q-values" (conditional values upon reaching the information set)~~

Problem: Opponent's strategy can change!

Consider this infoset for **P2**, and suppose that **P1** plays

(Exit, Rock) on 90% of timesteps

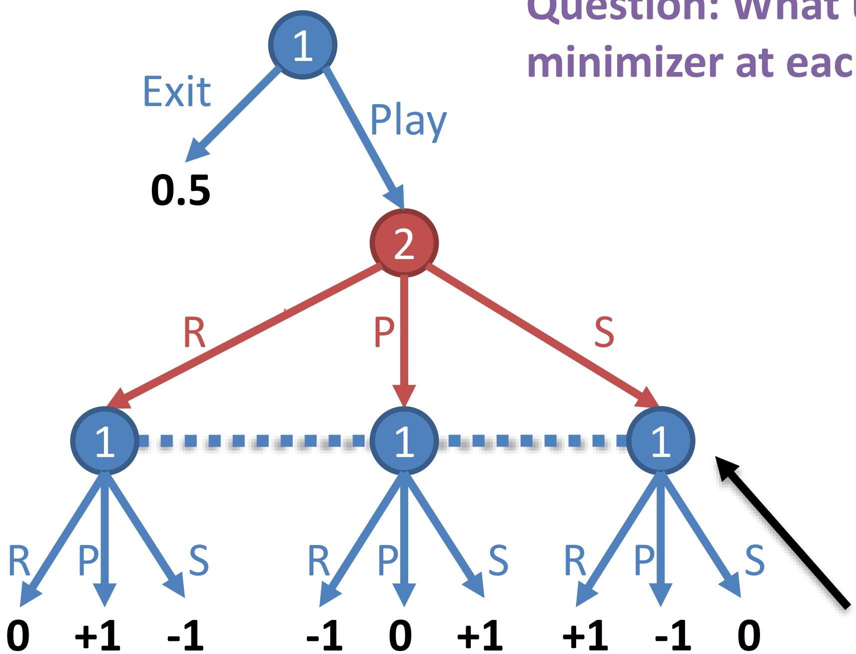
(Play, Paper) on 10% of timesteps

Regret minimizer receiving Q-value losses would play Paper, but it should play Scissors instead.

We should use some sort of "weighted" value!

Counterfactual Regret Minimization (CFR): The Gist

IDEA: Run one regret minimizer at each information set!



Question: What utility should we give to each infoset's regret minimizer at each time step?

~~**Attempt 2:** Use reach-weighted values:
utility of playing a at infoset I
= conditional value of a at I
* probability of reaching I~~

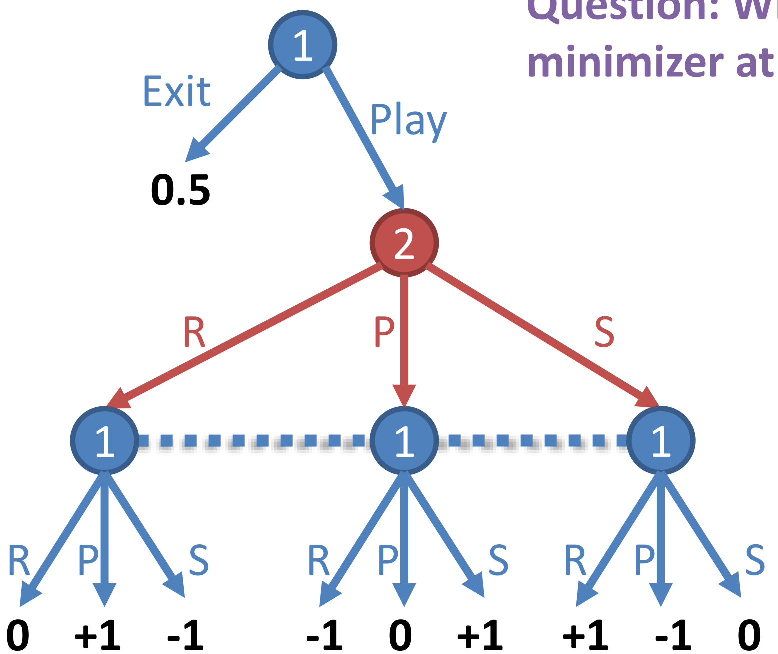
Problem: *Our* strategy can change *before* I

Suppose that **P2** always plays Rock
and **P1** is currently playing (Exit, Scissors)

This infoset isn't reached
⇒ its regret minimizer observes utility 0
⇒ **P1** never learns to play the correct best response (Play, Paper)!

Counterfactual Regret Minimization (CFR): The Gist

IDEA: Run one regret minimizer at each information set!



Question: What utility should we give to each infoset's regret minimizer at each time step?

Attempt 3: Use "counterfactual values":
utility of playing a at infoset I
= conditional value of I
* probability of *all **other** players* *(including chance)* reaching I

$$= \sum_{z \succcurlyeq Ia} x_i(\sigma_i(z) | Ia) \cdot x_{-i}\big(\sigma_{-i}(z)\big) \cdot c(z) \cdot u_i(z)$$

Pr[i plays all actions on $Ia \to \sigma_i(z)$ path]

THIS WORKS!

This is the algorithm called *counterfactual regret minimization (CFR)*

Proving the Correctness of CFR

Simple proof in this lecture due to G Farina, CK Ling, F Fang, T Sandholm (*NeurIPS* 2019), "Efficient Regret Minimization Algorithm for Extensive-Form Correlated Equilibrium" *e.g.*, FTRL, RM, RM+, …

Theorem: when using CFR in self-play in a 2p0s EFG with any regret minimizers whose regrets are bounded by $|A|\sqrt{T}$, the average strategy profile after T rounds is an ϵ-NE, where

$$\epsilon = \frac{|\Sigma_1| + |\Sigma_2|}{\sqrt{T}}$$

set of sequences

Scaled Extensions

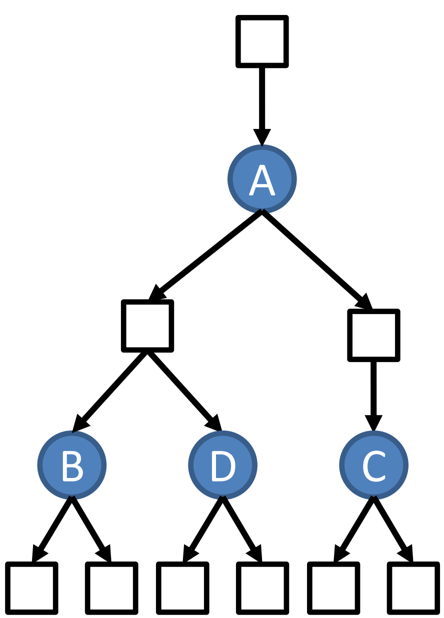
Idea: Construct a sequence-form strategy set **recursively**

Def: Given
- set $X \subseteq \mathbb{R}^n_{\geq 0}$
- vector $\boldsymbol{f} \in \mathbb{R}^n_{\geq 0}$ with $0 \leq \langle \boldsymbol{f}, \boldsymbol{x} \rangle \leq 1$ for all $\boldsymbol{x} \in X$
- integer $m > 0$

the **scaled extension** of X by the m-simplex under $\boldsymbol{f}$ is

$$X \lhd (\boldsymbol{f}, m) \coloneqq \left\{ \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{s} \end{pmatrix} \in X \times \mathbb{R}^m_{\geq 0} \;:\; \langle \boldsymbol{1}, \boldsymbol{s} \rangle = \langle \boldsymbol{f}, \boldsymbol{x} \rangle \right\}$$

$\boldsymbol{e}_\sigma$ = basis vector at coordinate σ



$$X = \{1\} \lhd (\boldsymbol{e}_\emptyset, 2) \lhd (\boldsymbol{e}_{\mathrm{A}\ell}, 2) \lhd (\boldsymbol{e}_{\mathrm{A}r}, 2) \lhd (\boldsymbol{e}_{\mathrm{A}\ell}, 2)$$

⋒

$\mathbb{R}$

Suffices (by induction) to construct regret minimizer on $X \lhd (f, m)$ given regret minimizer on X

CFR via Scaled Extensions

$$X \lhd (\boldsymbol{f}, m) := \left\{ \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{s} \end{pmatrix} \in X \times \mathbb{R}^m_{\ge 0} \;\; : \;\; \langle \mathbf{1}, \boldsymbol{s} \rangle = \langle \boldsymbol{f}, \boldsymbol{x} \rangle \right\}$$

$\mathcal{R}_X$: regret minimizer on $X \subseteq \mathbb{R}^n_{\ge 0}$

$\mathcal{R}_\Delta$: regret minimizer on Δ^m (e.g., MWU, RM, RM+, …)

Goal: Construct regret minimizer for $Y \;:=\; X \lhd (f, m)$

at each timestep t:

- $\boldsymbol{x}^t \leftarrow$ next strategy from $\mathcal{R}_X$
- $\boldsymbol{s}^t \leftarrow$ next strategy from $\mathcal{R}_\Delta$
- play $\boldsymbol{y}^t := (\boldsymbol{x}^t,\ \langle \boldsymbol{f}, \boldsymbol{x}^t \rangle \cdot \boldsymbol{s}^t) \in Y$
- receive utility $\boldsymbol{u}^t := \left(\boldsymbol{u}^t_X, \boldsymbol{u}^t_\Delta\right) \in \mathbb{R}^{n+m}$
- pass utility $\boldsymbol{u}^t_\Delta$ to $\mathcal{R}_\Delta$
- pass utility $\boldsymbol{u}^t_X + \left\langle \boldsymbol{u}^t_\Delta, \boldsymbol{s}^t \right\rangle \cdot \boldsymbol{f}$ to $\mathcal{R}_X$

Exercise: Check that this is equivalent to using counterfactual values (defined earlier).

CFR via Scaled Extensions

at each timestep t:

- $\boldsymbol{x}^t \leftarrow$ next strategy from $\mathcal{R}_X$
- $\boldsymbol{s}^t \leftarrow$ next strategy from $\mathcal{R}_\Delta$
- play $\boldsymbol{y}^t := (\boldsymbol{x}^t,\ \langle \boldsymbol{f}, \boldsymbol{x}^t \rangle \cdot \boldsymbol{s}^t) \in Y$
- receive utility $\boldsymbol{u}^t := \left(\boldsymbol{u}_X^t, \boldsymbol{u}_\Delta^t\right) \in \mathbb{R}^{n+m}$
- pass utility $\boldsymbol{u}_\Delta^t$ to $\mathcal{R}_\Delta$
- pass utility $\boldsymbol{u}_X^t + \left\langle \boldsymbol{u}_\Delta^t, \boldsymbol{s}^t \right\rangle \cdot \boldsymbol{f}$ to $\mathcal{R}_X$

$$X \lhd (\boldsymbol{f}, m) := \left\{ \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{s} \end{pmatrix} \in X \times \mathbb{R}_{\geq 0}^m \ : \ \langle \boldsymbol{1}, \boldsymbol{s} \rangle = \langle \boldsymbol{f}, \boldsymbol{x} \rangle \right\}$$

$$R_Y^T = \max_{\substack{\boldsymbol{x} \in X \\ \boldsymbol{s} \in \Delta^m}} \sum_{t=1}^T \left[\langle \boldsymbol{u}_X^t, \boldsymbol{x} \rangle + \left\langle \boldsymbol{u}_\Delta^t, \langle \boldsymbol{f}, \boldsymbol{x} \rangle \cdot \boldsymbol{s} \right\rangle - \langle \boldsymbol{u}_X^t, \boldsymbol{x}^t \rangle - \left\langle \boldsymbol{u}_\Delta^t, \langle \boldsymbol{f}, \boldsymbol{x}^t \rangle \cdot \boldsymbol{s}^t \right\rangle \right]$$

$$= \max_{\boldsymbol{x} \in X} \left[\sum_{t=1}^T \left[\langle \boldsymbol{u}_X^t, \boldsymbol{x} \rangle - \langle \boldsymbol{u}_X^t, \boldsymbol{x}^t \rangle - \left\langle \boldsymbol{u}_\Delta^t, \boldsymbol{s}^t \right\rangle \langle \boldsymbol{f}, \boldsymbol{x}^t \rangle \right] + \langle \boldsymbol{f}, \boldsymbol{x} \rangle \max_{\boldsymbol{s} \in \Delta^m} \sum_{t=1}^T \langle \boldsymbol{u}_\Delta^t, \boldsymbol{s} \rangle \right]$$

$$= \max_{\boldsymbol{x} \in X} \left[\sum_{t=1}^T \left[\langle \boldsymbol{u}_X^t, \boldsymbol{x} \rangle - \langle \boldsymbol{u}_X^t, \boldsymbol{x}^t \rangle - \left\langle \boldsymbol{u}_\Delta^t, \boldsymbol{s}^t \right\rangle \langle \boldsymbol{f}, \boldsymbol{x}^t \rangle \right] + \langle \boldsymbol{f}, \boldsymbol{x} \rangle \left(R_\Delta^T + \sum_{t=1}^T \langle \boldsymbol{u}_\Delta^t, \boldsymbol{s}^t \rangle \right) \right]$$

$$= \max_{\boldsymbol{x} \in X} \left[\underbrace{\sum_{t=1}^T \left[\langle \boldsymbol{u}_X^t, \boldsymbol{x} \rangle + \left\langle \boldsymbol{u}_\Delta^t, \boldsymbol{s}^t \right\rangle \langle \boldsymbol{f}, \boldsymbol{x} \rangle - \langle \boldsymbol{u}_X^t, \boldsymbol{x}^t \rangle - \left\langle \boldsymbol{u}_\Delta^t, \boldsymbol{s}^t \right\rangle \langle \boldsymbol{f}, \boldsymbol{x}^t \rangle \right]}_{R_X^T} + \langle \boldsymbol{f}, \boldsymbol{x} \rangle R_\Delta^T \right]$$

$$\leq R_X^T + \max_{\boldsymbol{x} \in X} \langle \boldsymbol{f}, \boldsymbol{x} \rangle R_\Delta^T \leq R_X^T + \left[R_\Delta^T \right]^+$$

CFR via Scaled Extensions

at each timestep t:

- $\boldsymbol{x}^t \leftarrow$ next strategy from $\mathcal{R}_X$
- $\boldsymbol{s}^t \leftarrow$ next strategy from $\mathcal{R}_\Delta$
- play $\boldsymbol{y}^t := (\boldsymbol{x}^t, \langle \boldsymbol{f}, \boldsymbol{x}^t \rangle \cdot \boldsymbol{s}^t) \in Y$
- receive utility $\boldsymbol{u}^t := \left(\boldsymbol{u}_X^t, \boldsymbol{u}_\Delta^t\right) \in \mathbb{R}^{n+m}$
- pass utility $\boldsymbol{u}_\Delta^t$ to $\mathcal{R}_\Delta$
- pass utility $\boldsymbol{u}_X^t + \langle \boldsymbol{u}_\Delta^t, \boldsymbol{s}^t \rangle \cdot \boldsymbol{f}$ to $\mathcal{R}_X$

$$X \lhd (\boldsymbol{f}, m) := \left\{ \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{s} \end{pmatrix} \in X \times \mathbb{R}^m_{\geq 0} \ : \ \langle \mathbf{1}, \boldsymbol{s} \rangle = \langle \boldsymbol{f}, \boldsymbol{x} \rangle \right\}$$

$$R_Y^T \leq R_X^T + \left[R_\Delta^T\right]^+$$

induction

with RM or RM+

set of sequences

CFR Regret Bound:

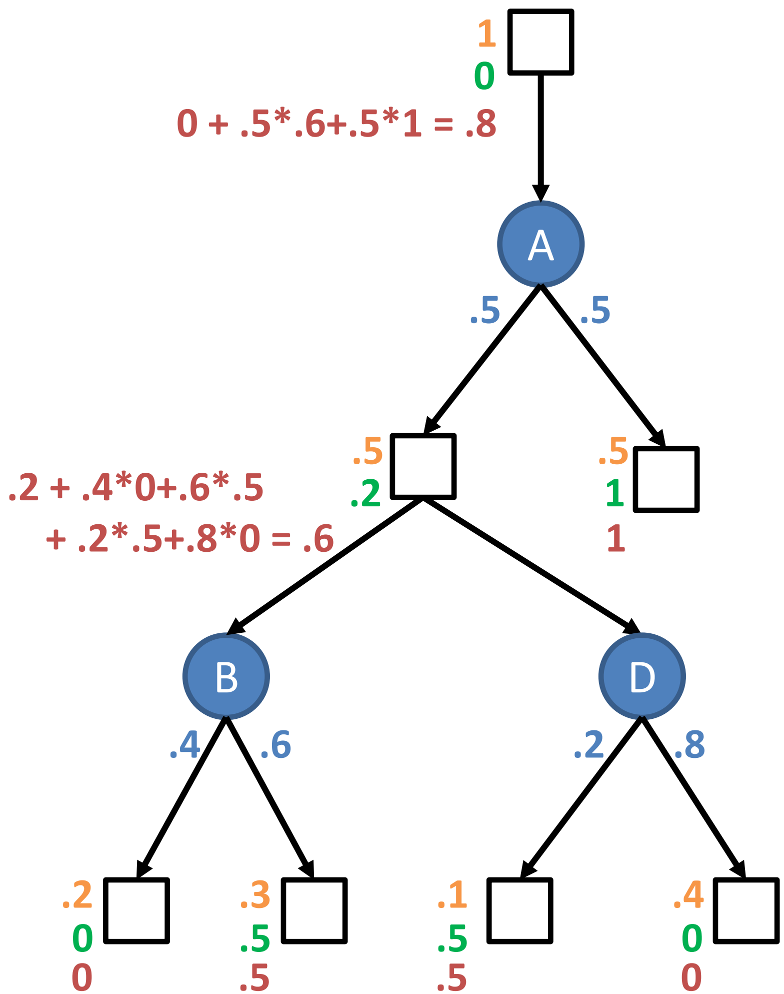
$$R_{X_i}^T \leq \sum_{I \in \mathcal{J}_i} [R_I^T]^+ \leq |\mathcal{J}_i| \cdot \sqrt{|A| T} \leq |\Sigma_i| \sqrt{T}$$

Theorem: when using CFR in self-play in a 2p0s EFG with any regret minimizers whose regrets are bounded by $|A|\sqrt{T}$, the average strategy profile after T rounds is an ϵ-NE, where

$$\epsilon = \frac{|\Sigma_1| + |\Sigma_2|}{\sqrt{T}}$$

□

Efficient Implementation



1. Query all local regret minimizers to get **behavioral strategies** π_1^t, π_2^t
2. Multiply down the tree to get **sequence-form strategies** x^t
3. Compute **utility vector** $\boldsymbol{u}^t := \boldsymbol{A}\boldsymbol{y}^t$
4. Compute **counterfactual values** (and pass them to local regret minimizers)

Time per iteration:
$O(\text{mul}(A) + |\Sigma_1| + |\Sigma_2|)$

time complexity of computing $\boldsymbol{A}\boldsymbol{y}^t$.
Trivially bounded by game tree size.

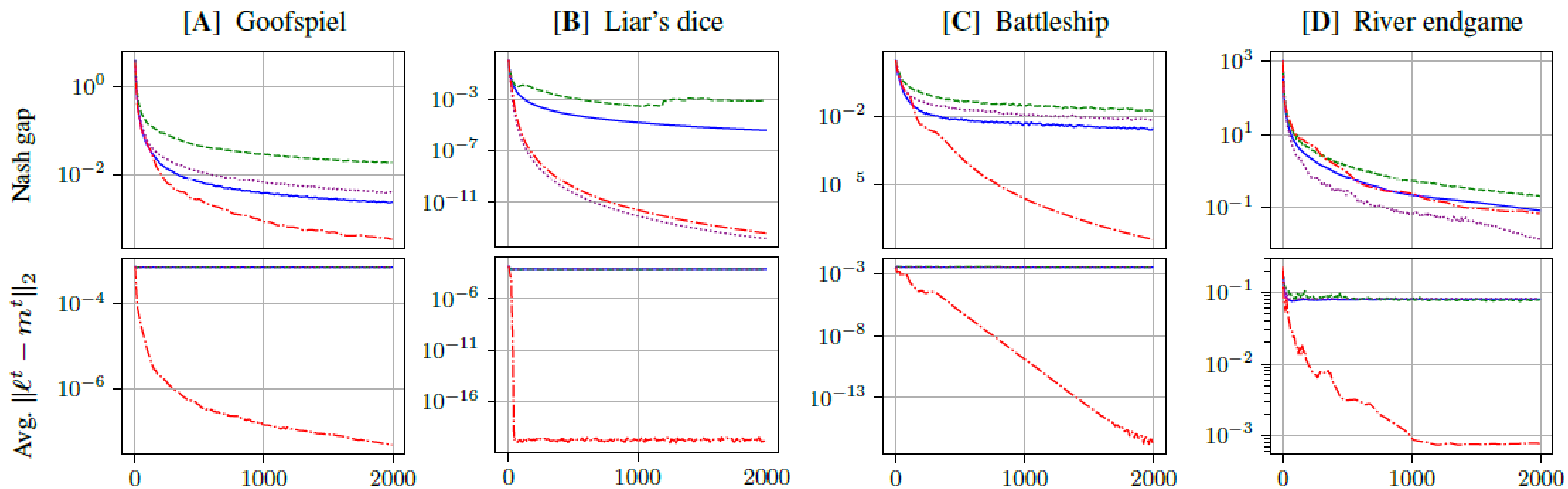
Exercise: Check that this actually implements CFR, *i.e.*, check that the values in Step 4 are actually the counterfactual values

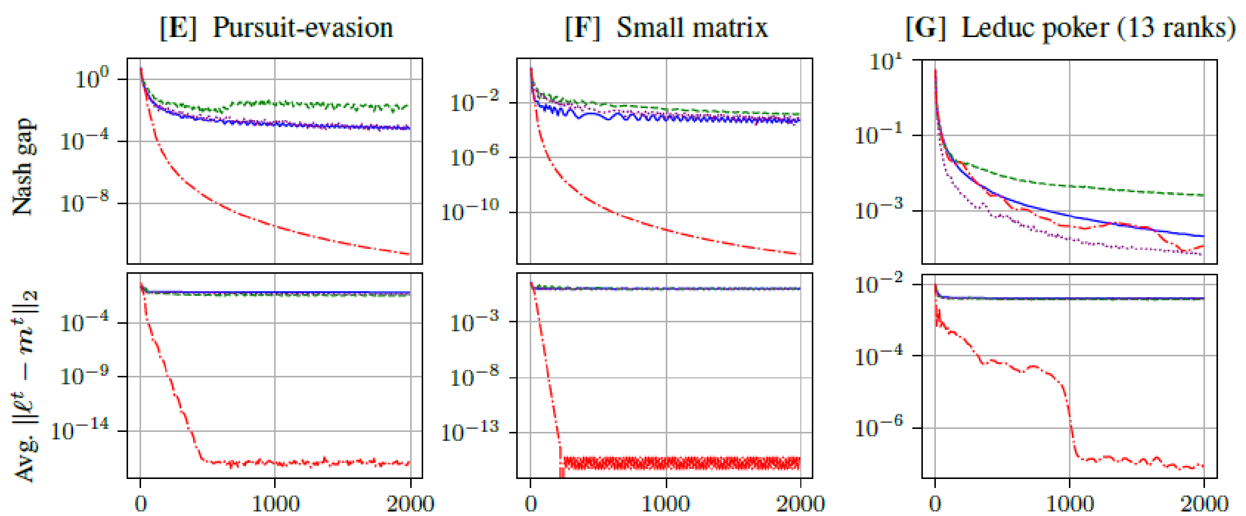
Why is CFR Superior in Practice?

- … to second-order methods (which can offer convergence rate $e^{-\Omega(T)}$)?
  - Does not require solving large linear systems
  - Second-order methods (interior point, …) don't fit in memory for large games

- … to general-purpose regret minimizers (e.g., FTRL)?
  - CFR uses an approach local to each decision point (easier to parallelize, warm-start, etc.) (*more on this next lecture!*)
    - [Brown & Sandholm, Reduced Space and Faster Convergence in Imperfect-Information Games via Pruning. ICML-17]
    - [Brown & Sandholm, Strategy-based warm starting for regret minimization in games, AAAI 2016]
  - No need for expensive projections onto feasible strategy polytope (think projected gradient descent)
  - **Scale-invariant!** (with RM/RM+)

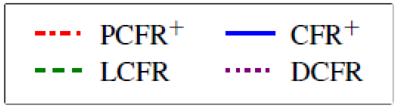
CFR Framework + Predictivity (aka optimism)





Dimension of the games

	Decision points	Actions	Leaves
[A]	4.4×10^6	5.3×10^6	1.7×10^6
[B]	2.5×10^4	4.9×10^4	1.5×10^5
[C]	1.1×10^6	3.2×10^6	3.5×10^6
[D]	4.4×10^4	1.3×10^5	5.4×10^7
[E]	1.2×10^4	6.9×10^4	1.2×10^5
[F]	2	4	4
[G]	5.1×10^3	1.2×10^4	9.9×10^4



Important Takeaways

- You can construct a regret minimizer for **sequential** decision making problems by combining regret minimizers for individual decision points
  - ⇒ Improvements on simplex domains carry over to extensive-form domains!

- Predictivity works well also in extensive-form domains

References

Kuhn's theorem:

- HW Kuhn (*Contrib. Theory of Games* 1950), "Extensive games and the problem of information"

Sequence form LP:

- B von Stengel (*GEB* 1996), "Efficient Computation of Behavior Strategies"

CFR:

- M Zinkevich, M Johanson, M Bowling, C Piccione (*NeurIPS* 2007), "Regret Minimization in Games with Incomplete Information“
- **Simple proof in this lecture due to** G Farina, CK Ling, F Fang, T Sandholm (*NeurIPS* 2019), “Efficient Regret Minimization Algorithm for Extensive-Form Correlated Equilibrium”