Solving Extensive-Form Games and Other Representations of Dynamic Games

Branislav Bošanský, Dominik Seitz, José Hilario , Michal Šustr

Czech Technical University in Prague

branislav.bosansky@agents.fel.cvut.cz
dominik.seitz@aic.fel.cvut.cz
hilarjos@fel.cvut.cz
michal.sustr@aic.fel.cvut.cz

November 12, 2019

Previously ... on multi-agent systems (tutorials and lectures).

1. Solving Extensive-Form Games
2. Sequence-Form Representation

Reminder: Sequence-form LP

Reminder from lectures:

$$\max_{r_1, v} v(root) \tag{1}$$

$$\text{s.t.} \quad r_1(\emptyset) = 1 \tag{2}$$

$$0 \leq r_1(\sigma_1) \leq 1 \qquad \forall \sigma_1 \in \Sigma_1 \tag{3}$$

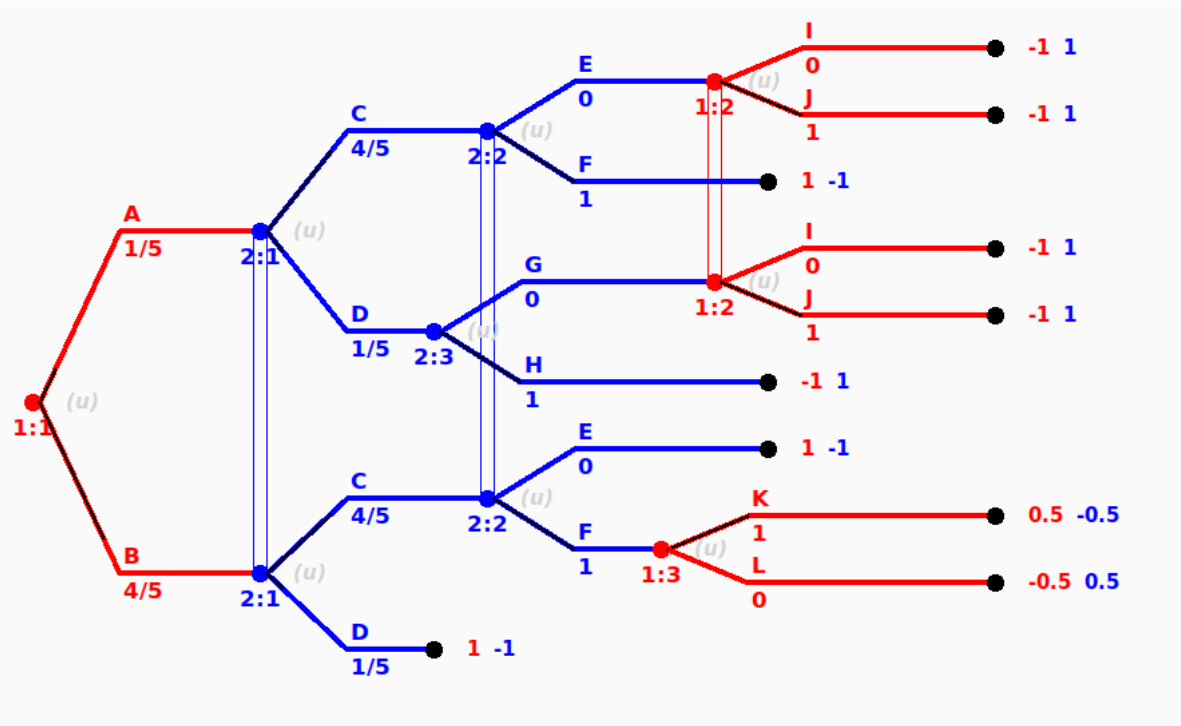
$$\sum_{a \in \mathcal{A}(I_1)} r_1(\sigma_1 a) = r_1(\sigma_1) \quad \forall I_1 \in \mathcal{I}_1, \sigma_1 = \mathsf{seq}_1(I_1) \tag{4}$$

$$\sum_{I' \in \mathcal{I}_2 : \sigma_2 a = \mathsf{seq}_2(I')} v(I') + \sum_{\sigma_1 \in \Sigma_1} g(\sigma_1, \sigma_2 a) r_1(\sigma_1) \geq v(I) \qquad \forall I \in \mathcal{I}_2, \sigma_2 = \mathsf{seq}_2(I), \forall a \in \mathcal{A}(I) \tag{5}$$

- $\mathsf{seq}_i(I)$ is a sequence of player i to information set,
- $I \in \mathcal{I}_i$, v_I is an expected utility in an information set,
- $\mathsf{inf}_i(\sigma_i)$ is an information set, where the last action of σ_i has been executed,
- $\sigma_i a$ denotes an extension of a sequence σ_i with action a

What happens if the a player moves twice in a row?

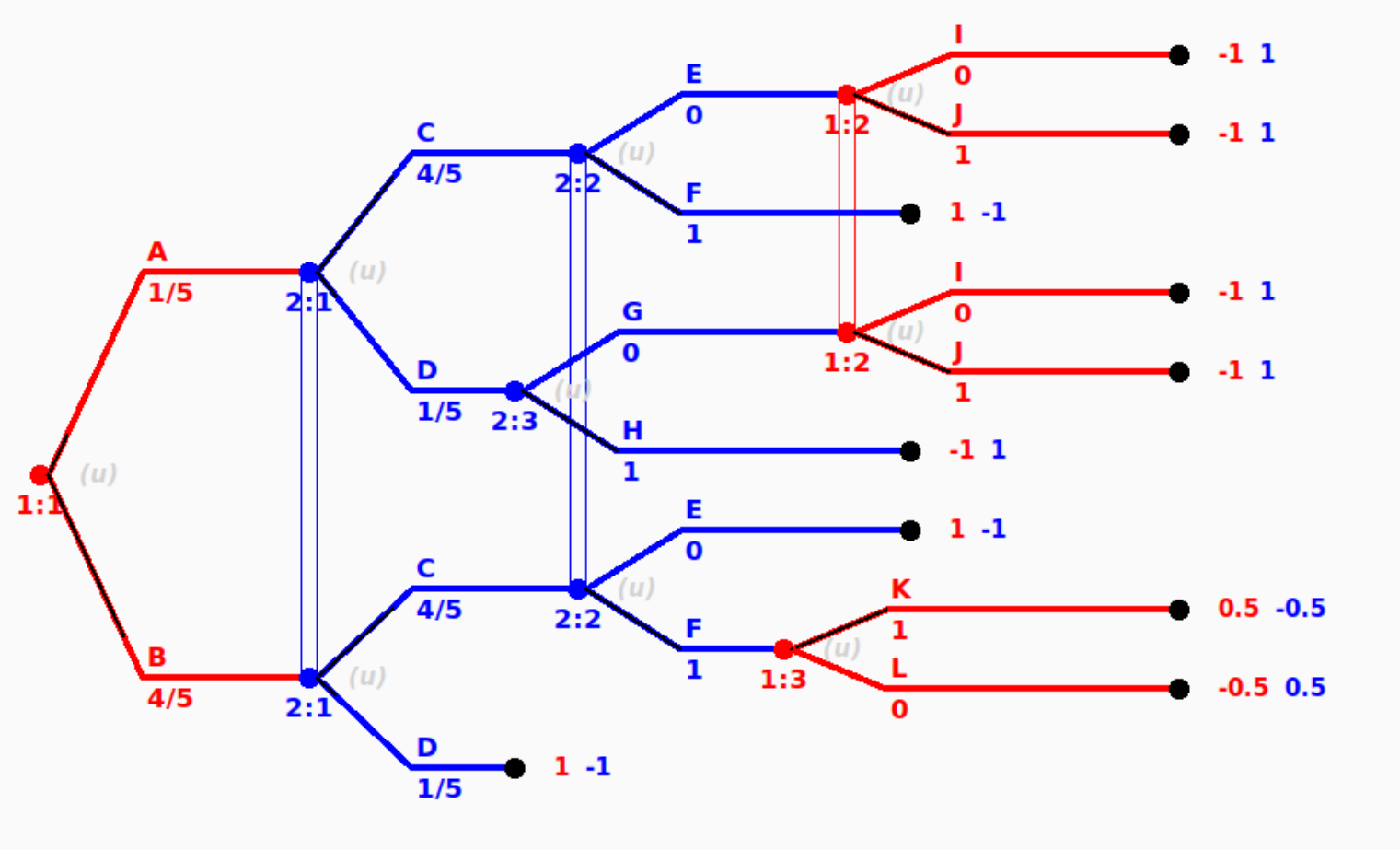
In the following EFG Player 2 (sometimes) moves twice in a row:



- How many infosets do each of the players have?
- How do the constraints change when constructing the sequence form LP for Player 2?

Sequence-Form LP Practice

Task 1: Write down a sequence-form linear program for both players:



Approximate Algorithms for Extensive-Form Games

- How else can we solve (large) EFGs?
- We can also learn the optimal strategy through self-play (See AlphaStar)
- An example of a self-play algorithm in game theory is Counterfactual Regret Minimization (CFR)

Main idea:

- in each iteration, traverse through the game tree and adapt the strategy in each information set according to the learning rule
- this learning rule minimizes the (counterfactual) regret
- the algorithm minimizes the overall regret in the game
- the average strategy converges to the optimal strategy

Reminder: Regret and Counterfactual Regret

Player i's regret for *not playing* an action a_i' against opponent's action a_{-i}

$$u_i(a_i', a_{-i}) - u_i(a_i, a_{-i})$$

In extensive-form games we need to evaluate the value for each action in an information set *(counterfactual value)*

$$v_i(s, I) = \sum_{z \in \mathcal{Z}_I} \pi^s_{-i}(z[I]) \pi^s_i(z|z[I]) u_i(z),$$

where

- $\mathcal{Z}_I$ are leafs reachable from information set I
- $z[I]$ is the history prefix of z in I
- $\pi^s_i(h)$ is the probability of player i reaching node h following strategy s

Reminder: Regret and Counterfactual Regret

Counterfactual value for one deviation in information set I; strategy s is altered in information set I by playing action $a : v_i(s_{I \to a}, I)$

at a time step t, the algorithm computes *counterfactual regret* for current strategy

$$r_i^t(I, a) = v_i(s_{I \to a}, I) - v_i(s_I, I)$$

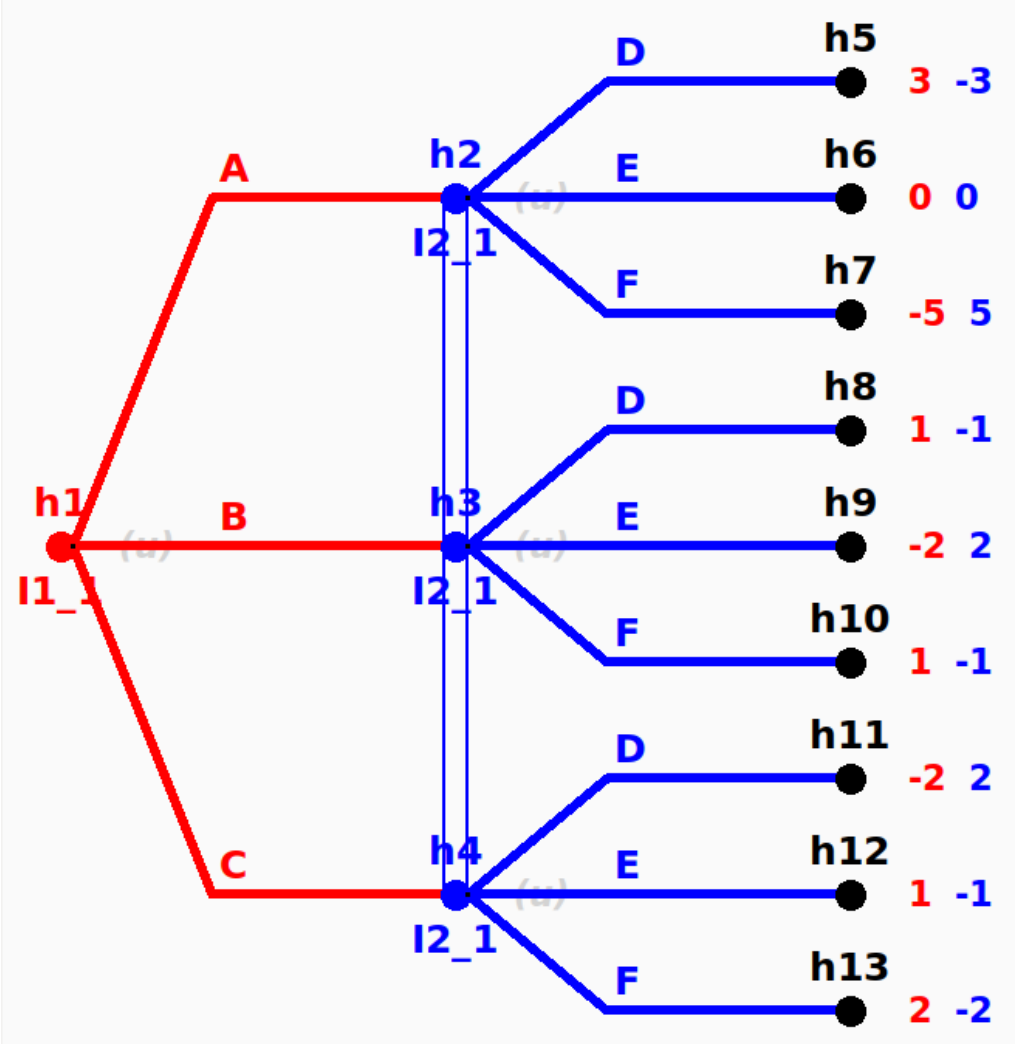
the algorithm calculates the *cumulative regret*

$$R_i^T = \sum_{t=1}^{T} r_i^t(I, a), \qquad R_i^{T,+}(I, a) = \max\{R_i^T(I, a), 0\}$$

strategy for the next iteration is selected using *regret matching*

$$s_i^{t+1}(I, a) = \begin{cases} \dfrac{R_i^{T,+}(I,a)}{\sum_{a' \in \mathcal{A}(I)} R_i^{T,+}(I,a')} & \text{if the denominator is positive} \\ \dfrac{1}{|\mathcal{A}(I)|} & \text{otherwise} \end{cases}$$

Solving an EFG using CFR - An Example

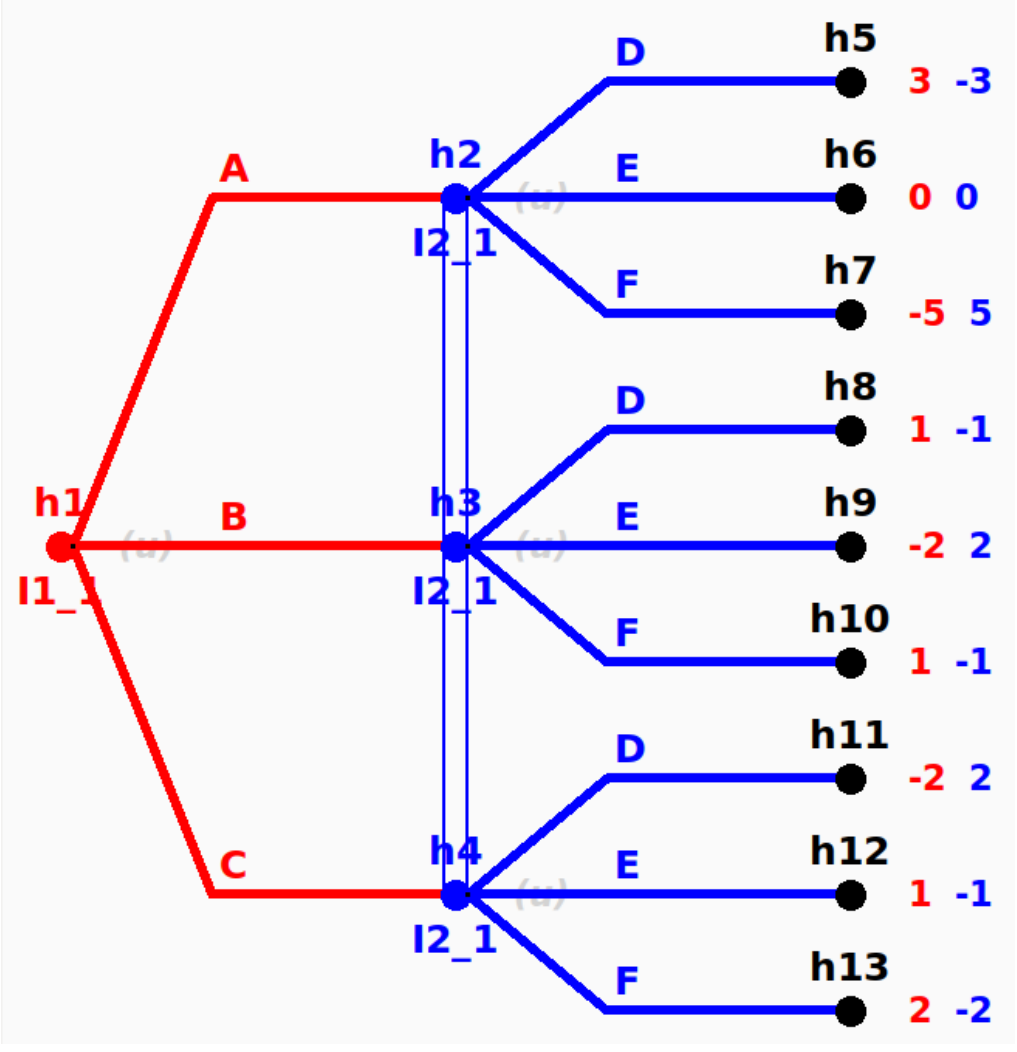


Solving an EFG using CFR - An Example

CFR needs to store regrets and cumulative average strategy for each infoset.

1. We have one infoset with three actions for each player.
2. We initialize a vector of regrets for each action of infoset $I_1 = A, B, C$ and $I_2 = D, E, F$
3. All actions are initially played with $\frac{1}{|\mathcal{A}(I)|}$ such that all actions are played with equal probability
4. We initialize an average strategy accumulator for each of them.
5. $R_{i=1}^{t=0}(I_1) = [0, 0, 0]$ and $R_{i=2}^{t=0}(I_2) = [0, 0, 0]$
6. $AvgAcc_{i=1}^{t=0}(I_1) = [0, 0, 0]$ and $AvgAcc_{i=2}^{t=0}(I_2) = [0, 0, 0]$

Task 2 - Run one iteration of CFR on this game



Task 2 - CFR Steps

In each iteration of CFR we compute:

1. Regrets for each action of each infoset $I_1 = A, B, C$ and $I_2 = D, E, F$
2. Multiply the regret by the reach probability of the opponent
3. Accumulate regret to $R_1^t(I_1)$ and $R_2^t(I_2)$
4. Multiply the probability of an action by the reach prob of the acting player
5. Accumulate it to $AvgAcc_1^t(I_1)$ and $AvgAcc_2^t(I_2)$

Example for counterfactual regret for action D at $h2$

1. Calculate expected value of node $h2$ given Player 2 plays uniformly ($u(h2)$)
2. Take difference between only playing D (going to $h5$ with prob 1) and $u(h2)$
3. Multiply by $\pi_1(h2)$ and accumulate to $r_2^t(D, I_2)$
4. Multiply the probability of an action D by the $\pi_2(h2)$
5. Accumulate it to $avgacc_2^t(D, I_2)$