CS 6840 Algorithmic Game Theory Dec 2, 2024

Lecture 38: Learning in extensive form games

Instructor: Eva Tardos *Scribe: Kiran Rokade*

1 Recap

Recall the notion of an extensive-form game. An extensive form game is described by a decision tree. The game begins at the root node. At each node, a fixed player makes a decision about which branch the game progresses on to, by either choosing a pure action, or by assigning probabilities on each action. At the end, when the game reaches a leaf node, the players earn payoffs. Each leaf node is assigned a fixed set of payoffs for all the players.

At each level of the game tree, a player may not know which node he is on. Such nodes are grouped together to form the player's "information sets". A player does not know which node within an information set he is on. However, he may know a distribution of probabilities on each node within the information set, by recalling the actions he played in the past, and having a belief on other players' strategies.

2 Strategy space and utilities for learning

We assume that the players have "perfect recall", i.e., they remember all the actions they chose in the past. At each information set, the player who makes a decision in the information set assigns probabilities to possible actions. Given an information set I, suppose player i makes a decision at this information set. For a node $v \in I$, let $x_v(a)$ be the probability that player i chooses action a when at node v. Since the player cannot distinguish between nodes within an information set, we impose the constraint $x_v(a) = x_w(a)$ for all $v, w \in I$. These probabilities form a strategy of the player. Let σ_i be the strategy of player i, and $\sigma = (\sigma_1, \ldots, \sigma_n)$ be the strategies of all the n players.

Next, we specify the utilities of each player. Let

$$u_i(v,\sigma) := \sum_{w \text{ leaf node below } v} \Pi^\sigma(v,w) u_i(w) \tag{1}$$

be the utility or value for player i in node v given the strategy σ, where $u_i(w)$ is the value of player i at the leaf node w, and $\Pi^\sigma(v,w)$ is the probability of reaching node w from node v given the strategy σ. However, recall that a player may not know which node he is on, instead he only knows which information set he is in, and knows the probability of being on a node within the information set. Let $\Pi^\sigma_{-i}(v)$ be the probability of reaching node v from the root node, given other players play using the strategy σ_{-i}, and player i makes decisions in favour of reaching v. Then,

$$u_i(I,\sigma) := \frac{\sum_{v \in I} \Pi^\sigma_{-i}(v) u_i(v,\sigma)}{\sum_{v \in I} \Pi^\sigma_{-i}(v)}. \tag{2}$$

is the value of player i in information set I under the strategy σ. This is also called the counterfactual value of player i conditioned on the player being in information set I.

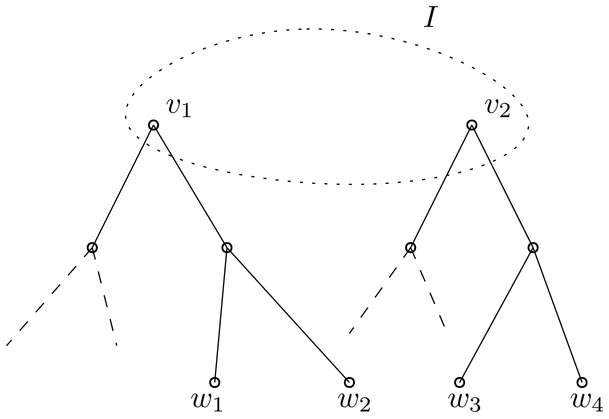

Figure 1: The value of information set I depends on the values of its nodes, which in turn depend on the values of the root nodes below them

3 Regret analysis

Suppose that players learn strategies by repeatedly playing the extensive form game and recursively update their strategies $\sigma^t = (\sigma_1^t, \dots, \sigma_n^t)$ at each time $t \in \{1, 2, \dots, T\}$. For this lecture, we assume that at each time instant t, all players know the strategies of all the other players. Let the instantaneous regret of player i at time t for not playing strategy A be

$$\mathrm{Reg}_i^t(A) := u_i(A, \sigma_{-i}^t) - u_i(\sigma^t) \tag{3}$$

where A is an alternative strategy of player i, and the total regret be

$$\mathrm{Reg}_i^T := \max_A \sum_{t=1}^T u_i(A, \sigma_{-i}^t) - u_i(\sigma^t). \tag{4}$$

The goal of the players is to learn a strategy that minimizes the total regret. Minimizing the total regret directly in an extensive form game can be challenging, since the size of the game tree can be very large for complicated games such as Poker. In this lecture, we show that the total regret can be minimized by simultaneusly minimizing the "immediate" or local regret across all nodes of the game tree. To this end, let the instantaneous immediate regret of player i at information set I at time t for not playing action a be defined as

$$\mathrm{Reg}_{i,\mathrm{im}}^t(I, a) := \Pi_{-i}^{\sigma^t}(I) \left[u_i(I, \sigma^t|_{I \to a}) - u_i(I, \sigma^t)\right] \tag{5}$$

where $\sigma^t|_{I\to a}$ denotes the strategy where player i switches to playing action a at information set I, while all other strategies remain unchanged. Let the total immediate regret over time T be

$$\mathrm{Reg}_{i,\mathrm{im}}^T(I) := \max_a \sum_{t=1}^T \Pi_{-i}^{\sigma^t}(I) \left[u_i(I, \sigma^t|_{I \to a}) - u_i(I, \sigma^t)\right]. \tag{6}$$

We will prove the following bound on the total regret in terms of the total immediate regret.

Theorem 1. *For all players i and time $T \geq 1$,*

$$\mathrm{Reg}_i^T \leq \sum_{I \textit{ info. sets of } i} \mathrm{Reg}_{i,\mathrm{im}}^T(I). \tag{7}$$

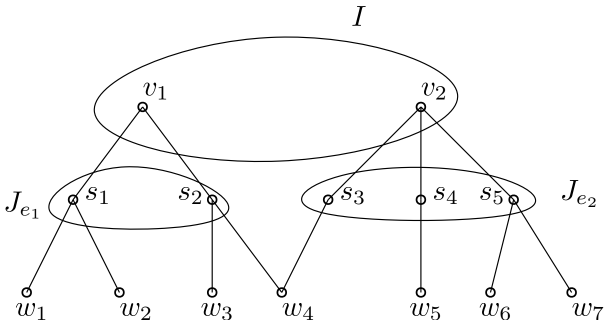

Figure 2: An illustration of the induction step

Thus, minimizing total immediate regret gives a bound on the total regret of player i. To prove the theorem, we prove the following more general claim.

Claim 1. *For all players i and time $T \geq 1$,*

$$\text{Reg}_i^T(I) \leq \sum_{J \textit{ info. sets of } i \textit{ under } I} \text{Reg}_{i,\text{im}}^T(J). \tag{8}$$

where

$$\text{Reg}_i^T(I) := \max_A \sum_{t=1}^{T} \Pi_{-i}^{\sigma^t}(I) \left[u_i(I, A, \sigma_{-i}^t) - u_i(I, \sigma^t)\right]. \tag{9}$$

is the total regret of player i starting from the information set I, and $u_i(I, A, \sigma_{-i}^t)$ is the utility of player i starting from information set I when player i plays strategy A at each time t while others play the strategy σ_{-i}^t.

If we can prove the claim, then Theorem 1 follows by applying the claim for I being the root node.

Proof of Claim 1: We prove the claim by induction on the height of the information set from the bottom of the tree.

For the base case, suppose I is the last information set of the player in the tree. Then, the claim holds trivially since the total immediate regret and the total regret are the same, hence both sides of the inquality are the same.

Otherwise, given an information set I, assume the induction hypothesis holds for all nodes below I, excluding I, i.e., for all information sets J_e directly below I,

$$\text{Reg}_i^T(J_e) \leq \sum_{K \text{ info. sets of } i \text{ under } J_e} \text{Reg}_{i,\text{im}}^T(K). \tag{10}$$

Let A be any fixed strategy of player i. Then, the instantaneous regret of player i for not playing the strategy A at time t is

$$\begin{aligned} \text{Reg}_i^t(I, A, \sigma) &:= \Pi_{-i}^{\sigma^t}(I)\left[u_i(I, A, \sigma_{-i}^t) - u_i(\sigma^t)\right] & (11) \\ &= \underbrace{\Pi_{-i}^{\sigma^t}(I)\left[u_i(I, A, \sigma_{-i}^t) - u_i(I, \sigma^t|_{I\to a})\right]}_{\text{Term 1}} + \underbrace{\Pi_{-i}^{\sigma^t}(I)\left[u_i(I, \sigma^t|_{I\to a}) - u_i(\sigma^t)\right]}_{\text{Term 2}} & (12) \end{aligned}$$

where a is the action of strategy A of player i at information set I.

First, note that

$$\text{Term } 2 = \text{Reg}^t_{i,\text{im}}(I, a). \tag{13}$$

Next, we bound Term 1. Note that

$$\begin{aligned}\Pi^{\sigma^t}_{-i}(I) u_i(I, A, \sigma^t_{-i}) &= \sum_{v \in I} u_i(v, A, \sigma^t_{-i}) \Pi^{\sigma^t}_{-i}(v) && (14)\\ &= \sum_{v \in I} \Pi^{\sigma^t}_{-i}(v) \sum_{\substack{w \text{ leaf below } v\\ \text{reachable by } A}} u_i(w) \Pi^{\sigma^t}_{-i,A}(v, w), && (15)\end{aligned}$$

where we used $\Pi^{\sigma^t}_{-i,A}(v, w)$ to denote the probability of reaching w from v with distribution σ_{-i} and i using the strategy A. Note that each (v, w) path goes through a node s which is in one of the information sets J_e directly below I (see Figure 2). Thus,

$$\begin{aligned}&\sum_{v \in I} \Pi^{\sigma^t}_{-i}(v) \sum_{\substack{w \text{ leaf below } v\\ \text{reachable by } A}} u_i(w) \Pi^{\sigma^t}_{-i,A}(v, w) && (16)\\ &= \sum_e \sum_{\substack{s \in J_e\\ \text{reachable by } a}} \Pi^{\sigma^t}_{-i}(s) \underbrace{\sum_{w \text{ leaf below } s} u_i(w) \Pi^{\sigma^t}_{-i,A_e}(s, w)}_{u_i(s, A_e, \sigma^t_{-i})} && (17)\\ &= \sum_e \sum_{\substack{s \in J_e\\ \text{reachable by } a}} \Pi^{\sigma^t}_{-i}(s) u_i(s, A_e, \sigma^t_{-i}) && (18)\end{aligned}$$

where A_e is the part of strategy A played from J_e onwards. Similarly,

$$\Pi^{\sigma^t}_{-i}(I) u_i(I, \sigma^t|_{I \to a}) = \sum_e \sum_{s \in J_e} \Pi^{\sigma^t}_{-i}(s) u_i(s, \sigma^t) \tag{19}$$

Hence,

$$\text{Term } 1 = \sum_e \sum_{s \in J_e} \Pi^{\sigma^t}_{-i}(s) \left[u_i(s, A_e, \sigma^t_{-i}) - u_i(s, \sigma^t) \right]. \tag{20}$$

Substituting the bounds derived above on Term 1 and Term 2 in equation (12) and summing over all t,

$$\begin{aligned}\sum_{t=1}^T \text{Reg}^t_i(I, A, \sigma) &\le \sum_{t=1}^T \sum_e \sum_{s \in J_e} \Pi^{\sigma^t}_{-i}(s) \left[u_i(s, A_e, \sigma^t_{-i}) - u_i(s, \sigma^t) \right] + \sum_{t=1}^T \text{Reg}^t_{i,\text{im}}(I, a) && (21)\\ &= \sum_e \underbrace{\sum_{t=1}^T \sum_{s \in J_e} \Pi^{\sigma^t}_{-i}(s) \left[u_i(s, A_e, \sigma^t_{-i}) - u_i(s, \sigma^t) \right]}_{\le \text{Reg}^T_i(J_e)} + \underbrace{\sum_{t=1}^T \text{Reg}^t_{i,\text{im}}(I, a)}_{\le \text{Reg}^T_{i,\text{im}}(I)} && (22)\\ &\le \sum_e \sum_{K \text{ info. sets of } i \text{ under } J_e} \text{Reg}^T_{i,\text{im}}(K) + \text{Reg}^T_{i,\text{im}}(I) && (23)\\ &= \sum_{J \text{ info. sets of } i \text{ under } I} \text{Reg}^T_{i,\text{im}}(J). && (24)\end{aligned}$$

where the last inequality is by the induction hypothesis. This completes the proof of the claim.

□