Computational Learning Theory **Spring Semester, 2005/6**

Lecture 7: May 9

Lecturer: Yishay Mansour Scribe: Eyal Schneider, Zack Dvey-Aharon, Sharon Bruckner

[1]

7.1 Extensive Games with Perfect Information

An extensive game is a detailed description of the sequential structure of the decision problems, encountered by the players in strategic situation.
There is perfect information in such a game if each player, when making any decision, is perfectly informed of all the events that have previously occurred.

7.1.1 Definitions

Definition An **extensive game with perfect information** $\langle N, H, P, U_i \rangle$ has the following components:

- A set of N players
- A set H of sequences (finite or infinite):
  H^t : The set of all possible histories up to time t.
  $H = \bigcup H^t$
  Each component of a history is an **action** taken by a player.
- P is the **player function**, $P(h)$ being the player who takes an action after the history h.
- Payoff function for player i is $U_i : H \to R$

After any history h player $P(h)$ chooses an action from the set $A(h) = \{a : (h, a) \in H\}$. The empty history is the starting point of the game. The game must end eventually. In other words, the length of all items in H is bounded.

[1]These notes are adapted from the scribe notes of Gur Yaari and Idan Szpektor, 2004

Example 1

Splitting a cake:

Two players try to split a cake in such a way that both are satisfied (see figure 7.1). Player 1 splits to two parts, and then player 2 chooses his part.

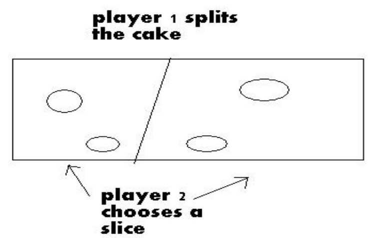

Figure 7.1: Splitting a cake, satisfying both parties

The one that split the cake is satisfied, because she believes the two pieces are equal. The second party cannot complain, since she chose her piece.

Example 2

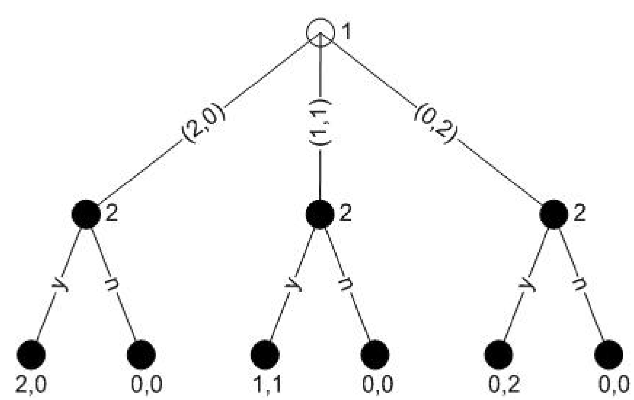

Figure 7.2: An extensive game, allocating two identical objects between two people

Two identical objects are to be given to two people. One of them proposes an allocation which the other then either accepts or rejects. They are both reasonable.
In this representation each node corresponds to a history and any edge corresponds to an action.

- $H = \{\emptyset, (2,0), (1,1), (0,2), ((0,2), y), ((2,0), n), ((1,1), y), ((1,1), n), ((0,2), y), ((0,2), n)\}$
- $P(\emptyset) = 1$ and $P(h) = 2, h \neq \emptyset$

Example 3

Defining a negotiation game:
Each player offers a price in her turn:

$P(h) = (|h| \ mod \ 2) + 1$
Player 1 offers a price, player 2 offers a price, and so on periodically.

Outcome:
Success: Equal prices.
Failure: No success after t rounds.

7.1.2 Strategy

Definition A **strategy of player** $i \in N$ in an extensive game $\langle N, H, P, \{U_i\}\rangle$ is a function that assigns an action in $A(h)$ to each history $h \in H$ for which $P(h) = i$

A strategy specifies the action chosen by a player for *every* history after which it is her turn to move, *even for histories that , if the strategy is followed, cannot be reached.*

Example

$S_1 = \{AE, AF, BE, BF\}$ - her strategy specifies an action after the history (A, C), even if she chooses B at the beginning of the game.
One can transform an extensive game with perfect information to a normal game by setting all the possible histories as the possible choices for a normal game.

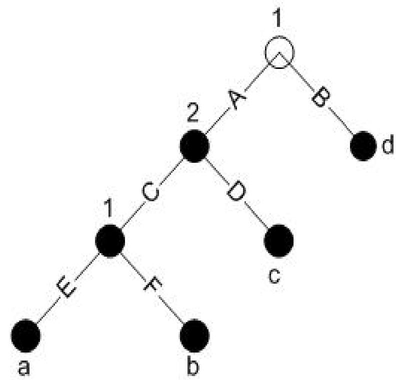

Figure 7.3: An extensive game in which player 1 moves before and after player 2

7.1.3 Nash Equilibrium

Definition A **Nash equilibrium of an extensive game with perfect information** $\langle N, H, P, \{U_i\}\rangle$ is a strategy profile $s^* = (s_i)_{i\in N}$ such that for every player $i \in N$ and for every strategy s we have $U_i(s^*) \geq U_i(s)$

Example

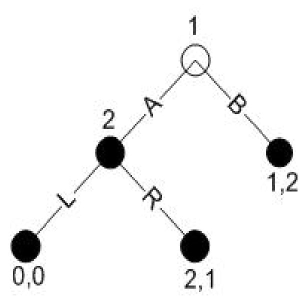

Figure 7.4: Two players extensive game

The game has two Nash equilibrium strategies: (A, R) and (B, L) with payoff $(2, 1)$ and $(1, 2)$ respectively. The strategy profile (B, L) is a Nash equilibrium because given that player 2 chooses L, it is optimal for player 1 to choose B at the beginning. The strategy (B, R) is not a Nash equilibrium since in that case player 1 prefers action A. Player 2's choice L is a "threat" for player 1, in order to motivate player 1 not to choose A. If player 2 chooses R, then player 1 prefers A since her payoff increases.

7.1.4 Subgame perfect Equilibrium

Definition A **subgame of the extensive game with perfect information** $\Gamma = \langle N, H, P, \{U_i\}\rangle$ that follows after the history h is the extensive game $\Gamma(h) = \langle N, H|_h, P|_h, \{U_i|_h\}\rangle$ where:
$H|_h$ is the set of sequences h' of actions for which $(h, h') \in H$
$U_i|_h : H|_h \to R$ and
$P|_h : H|_h \to N$

Definition A **subgame perfect equilibrium of an extensive game with perfect information** $\langle N, H, P, \{U_i\}\rangle$ is a strategy profile s^* such that for every player $i \in N$ and every history $h \in H$ for which $P(h) = i$ we have $U_i(s^*|_h) \geq U_i(s|_h)$ for every strategy s_i of player i in the subgame $\Gamma(h)$.

Lemma 7.1 *The strategy profile s^* is a subgame perfect equilibrium if and only if for every player $i \in N$ and every history $h \in H$ for which $P(h) = i$ and for every $a_i \in A_i(h)$, the following holds: $U_i(s^*|_h) \geq U_i(s^*_{-i}|_h, s_i)$ such that s_i differs from $s^*_i|_h$ only in the action a_i after the history h.*

Proof:
$\Rightarrow$ If s^* is a subgame perfect equilibrium then it is resilient to any deviation.
$\Leftarrow$ Suppose s^* is not an S.G.P equilibrium. Then exists a history h in which player $P(h)$ prefers to change her strategy. Let h be the longest history as above. For $P(h) = i$ she can change to some action $a_i \in A_i(h)$ and increase her payoff. Therefore there exists a single action as presented in the lemma. □

Theorem 7.2 *Every extensive game with perfect information has a subgame perfect equilibrium.*

Proof: We will use a backwards induction procedure. We start from the leaves and walk up through the tree layer by layer. For each vertex located at a path determined by history h, player $P(h)$ chooses her best action (Best Response). By lemma 7.1 this profile is a subgame perfect equilibrium. □

7.2 Repeated Games

The idea behind repeated games is that if we let the players play the same game multiple times, they might obtain different equilibria than those of a single game. Unlike the extensive form game, here the players choose actions concurrently. For example, we would like to achieve cooperation in the Prisoner's Dilemma game.

7.2.1 Finitely Repeated Games

Consider again the Prisoner's Dilemma game:

	C	D
C	$(3,3)$	$(0,4)$
D	$(4,0)$	$(1,1)$

Claim 7.3 *In a repeated game of T steps, the only Nash Equilibrium is to play (D,D) in all T steps.*

Proof: In the last step, both players will play D, since otherwise at least one of the players would want to change her decision in order to improve her payoff (one can see (D, D) is a dominant action). Using induction, if both players played (D, D) in the last i steps, the same reason holds for the $i - 1$ step. □

Consider a modified Prisoner's Dilemma game:

	C	D	E
C	$(3,3)$	$(0,4)$	$(0,0)$
D	$(4,0)$	$(1,1)$	$(0,0)$
E	$(0,0)$	$(0,0)$	$(\frac{1}{2},\frac{1}{2})$

Claim 7.4 *In the finite T steps modified Prisoner's Dilemma game, there is a subgame perfect equilibrium for which the outcome is (C, C) in every step but the last three, in which it is (D, D).*

Proof: The strategy of each player would be to play $T - 3$ times C and then play D the last 3 times. However, if the other player plays differently than this strategy, the player will play E for the rest of the steps. Since the cooperation stops at the $T - 2$ step, it's enough to consider the case in which player 2 has played differently at the $T - 3$ step (otherwise her payoff is even smaller). Here are the two possible outcomes starting from the $T - 3$ step:
Playing according to the strategy yields (C, C) (D, D) (D, D) (D, D). The total payoff of player 2 is $3 + 1 + 1 + 1 = 6$.
If the second player changes her strategy, the best moves that can be made are (C, D) (E, E)

(E,E) (E,E). The total payoff of the player 2 is $4+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}=5\frac{1}{2}$.
One can see that playing differently than the stated strategy yields less profit for the deviating player. Thus it is best to play the proposed strategy by both players. □

The average payoff in this game is $(3(T-3)+3)/T$ which is $3-\frac{6}{T}$. This payoff is close to 3 which is the payoff of repeated cooperation.

7.2.2 Infinitely Repeated Games

An **infinitely repeated game** is a game played an infinite number of times.

There are several ways to look at the payoff of a player in an infinitely repeated game. Consider a N-player game G with a payoff function $\vec{u}$, where u^i is the payoff function of player i. We define u^i_t as the payoff of player i at step t.

Definition The **average payoff** of a game G is the limit of the average payoff of the first T steps:

$$\frac{1}{T}(\sum_{t=1}^{T} u^i_t) \underset{T\to\infty}{\longrightarrow} \overline{u}^i$$

Definition The **finite payoff** of a game G is the sum of the payoffs of the first H steps of the game: $\sum_{t=1}^{H} u^i_t$

Definition The **discount payoff** of a game G is the weighted sum of the payoffs of the steps of the game: $\sum_{t=1}^{\infty} \gamma^t u^i_t$, where $0<\gamma<1$.

From this point onwards, the payoff of an infinitely repeated game will refer to the average payoff $\overline{u}^i$.

Definition The **payoff profile** of an infinitely repeated game G is the payoff vector $\vec{w}$, where w_i is the payoff of player i. A payoff profile $\vec{w}$ is *feasible* if there are β_a for each outcome $a\in A$, $K=\sum_{a\in A}\beta_a$, such that $\vec{w}=\sum_{a\in A}\frac{\beta_a}{K}\vec{u}(a)$.

Definition The **minimax payoff** of player i in a single step is: $v_i = \min_{a^{-i}\in A^{-i}} \max_{a_i\in A_i} u^i(a^{-i},a^i)$

Claim 7.5 *In every pure Nash Equilibrium of a single game, the payoff of player i is at least* v_i.

Proof: If a player has a smaller payoff than v_i then by the definition of the minimax payoff, there is a different strategy in which she can profit at least v_i, regardless of the other player's action a^{-i} □

Definition A payoff profile $\vec{w}$ is **enforceable** if $\forall_{i\in N}\ v_i \le w_i$. A payoff profile is **strictly enforceable** if $\forall_{i\in N}\ v_i < w_i$.

Theorem 7.6 *For every feasible enforceable payoff profile $\vec{w}$ of a given infinitely repeated game G, there exists a shared pure strategy s which is a pure Nash Equilibrium having $\vec{w}$ as its average payoff.*

Proof: We will describe a strategy that is Nash Equilibrium with the payoff $\vec{w}$.
Since $\vec{w}$ is feasible there are β_a for each $a \in A$, $K = \sum_{a \in A} \beta_a$, such that $\vec{w} = \sum_{a \in A} \frac{\beta_a}{K} \vec{u}(a)$.
We shall assume that $\forall_{a \in A} \beta_a \in \mathbf{N}$.
The strategy of each player is to play cycles of K steps, going over all the possible outcomes $a \in A$ in an ordered list and playing her outcome in a β_a times. If player i deviates from this strategy, the rest of the players will change to a new strategy P_{-i} that enforces the payoff of player i to be at most the minimax payoff v_i.
Thus, if a player i deviates from the main strategy, her payoff will be v_i. which is not better than her payoff in $\vec{w}$. Because each deviation will not improve the payoff of player i, $\vec{w}$ is a Nash Equilibrium. □

Theorem 7.7 *Every feasible strictly enforceable payoff profile $\vec{w}$ in an infinitely repeated game G has a Subgame Perfect pure Equilibrium with an average payoff $\vec{w}$.*

Proof: We will describe a strategy that is Subgame Perfect Equilibrium with the payoff $\vec{w}$.
We shall use the same cyclic strategy as in the previous theorem, where all the players play the outcome a for β_a steps. If a player deviates from the strategy, the other players will punish her but only in a finite number of steps. At the end of the punishing steps all players will resume to play the cyclic strategy.
More specifically, at the end of each K steps cycle, the players will check if one of the players has deviated from the cyclic strategy. If a player, let say player j, has indeed played differently, the other players will play, for m^* steps, the minimax strategy P_{-j} that will enforce a payoff of at most v_j for player j, where m^* is chosen as follows:
We mark player j's strategy in each step t of the last cycle as $a_j^{'t}$. The maximal payoff benefit for player j out of the steps in the cycle is $g^* = \max_{a^t}[u^j(a_{-j}^t, a_j^{'t}) - u^j(a^t)]$, where $a^t \in A$ is the expected outcome in step t of the K steps. We would like to get $Kg^* + m^* v_j < m^* w_j$ in order to make the punishment worthwhile. However, since $\vec{w}$ is strictly enforceable, we know that $v_j < w_j$ and so there exist m^* such that $0 < \frac{Kg^*}{w_j - v_j} < m^*$.
Playing the punishment strategy for m^* steps will yield a strictly smaller payoff for player j than playing the cyclic strategy without deviation. □

7.3 Bounded Rationality

We have seen that in a repeated Prisoner's Dilemma game of N rounds there is no cooperation. A way to circumvent this problem is to assume that the players have limited resources. This kind of player is said to have **Bounded Rationality**.

A Bounded Rationality player i is an automaton defined as follows:

S - the states set.

A - the action vectors (possible game outcomes).

$\delta : S \times A \rightarrow S$ - the state transitions.

$f : S \rightarrow A_i$ - the actions function.

S_0 - the initial state.

We assume that the automaton is deterministic and that in each state only one action is chosen. A stochastic strategy is to randomly choose one automaton from a set of deterministic automata.

7.3.1 Tit for Tat

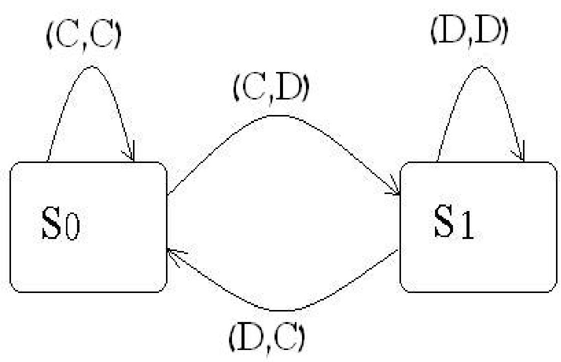

Figure 7.5: The automata of the Tit for Tat strategy

The Tit for Tat strategy (TfT) for the repeated Prisoner's Dilemma game consists of playing what the opponent played in the last round (see Figure 7.5).

Theorem 7.8 *Consider a N round Prisoner's Dilemma game between two automata. If the automata have at most $N-1$ states then (TfT, TfT) is a Nash-Equilibrium.*

Proof:
Suppose that player 2 played $N-1$ times C. Then its automaton must have returned, at some point, to an already visited state. This will create a loop, since the input from player 1 continues to be C. That means that player 2 will also play C in the last round.

Now suppose that the first time that player 2 diverges from the cooperative strategy happens during k consecutive rounds, which are not the last ones of the game. During these rounds and the following one, its payoff will be $4 + (k-1) * 1 + 0 = k + 3$. Alternatively, cooperating during these $k+1$ rounds would result in a payoff of $3k+3$, which is better. In the case of not cooperating during the last k rounds of the game, we have a payoff of $k+3$ against $3k$, which is profitable only when $k = 1$. However, as we proved before, player 2 will never be able to defect the first time in the last round.

Thus, if both players follow the TfT strategy, the system is in Nash-Equilibrium. □

7.3.2 Bounded Prisoner's Dilemma

We have seen an example for a simple bounded rationality strategy for the Prisoner's Dilemma game that will yield a cooperation in some conditions. The next step is to analyze any general bounded rationality strategy for that game, described by a final automaton, and find the conditions that will lead to a cooperation between two players using these strategies in an N round game.

Theorem 7.9 *In a Repeated Prisoner's Dilemma with N rounds, where both players have an automaton with at least 2^{N-1} states, the only Equilibrium is the one in which both players play (D, D) in all rounds.*

Proof: We will prove that using the 2^{N-1} states, any strategy can be coded. The reduction from this case of Bounded-Rationality to a general strategy, proves that the equilibrium is identical to the general case, i.e. both players play always (D, D).

Let S be a general strategy to be coded. We assume that the Nth play of S is D, otherwise S is not optimal.

The automaton based on S consists of a binary tree of height $N-2$. Each node(state) represents a possible history in the game, determined by its path from the root. Each node is assigned an action according to the strategy S. Note that the tree layer at depth $N-1$ is not necessary, since S specifies D as the Nth action, regardless of the history. Instead, a single node will be used. This node represents the state after $N-1$ actions.

The total number of nodes in this tree is 2^{N-1}.

□

A much more involved theorem is the following, showing that in case the number of states is sub-exponential, cooperation can occur.

Theorem 7.10 *For a Repeated Prisoner's Dilemma with N rounds, when both players have automata with at most $2^{\epsilon_1 N}$ states (when it is possible to change to a different automata with a related size boundary $2^{\epsilon_2 N}$), there exists an Equilibrium with a profit of $3 - \epsilon_3$.*