Computer Science 497 SML
Transcription of the class notes for March 4 and March 6, 2003

March 31, 2003

Schedule

Recently Covered

Imperfect State Information

- Probabilistic Information Spaces
- Nondeterministic Information Spaces

Covered this week

Game Theory

- Overview
- Zero-Sum Games
  - Saddle Points
  - Mixed-Strategies
  - Computing Equillibria
- Non-zero Sum Games

Yet to be covered

Sequential Games

Game Theory

Overview

Game theory is the field that explains the decision theoretic behavior of multiple decision makers, which are traditionally called players in game theory literature. One book on game theory, kept on reserve for this class, is "Dynamic Cooperative Game Theory" by Tamer Basar and Geert Jan Olsder, 2nd edition.

There are many variations within game theory.

- Single Stage vs. Multiple Stage: The question here is "Will you ever have the opportunity to play against the same player again?" If not, you might as well play to kill. If so, you might be nice to try to coax them into moving in a mutually favorable cooperative direction.
- Different Information Spaces for both Players: Do players have a different information about the state space, or even about the nature of the game? One classic game with unknown information is Battleship, in which each player cannot see each other's pieces, and attempts to learn where those pieces are.

- Zero Sum vs. Non-zero Sum: Are you strictly opposed (my gain is your loss) or is there some asymmetry in which there are cooperative actions even if you pursue your own interests.
- Deterministic vs. Randomized Strategies (called pure vs. mixed in game theory literature)
- Cooperative vs. Non-cooperative: In games with multiple stages, especially in games with more than two players, it is quite productive to have cooperative elements to your strategy. In economics, if you can form a cartel, then you effectively create a monopoly, determining the price instead of being subject to the market. In reality TV shows with voting, like Survivor, it behooves you to form collaborations.

One common assumption that is made is that each player knows all of the other players' cost functionals. This is not necessarily the case. So, even the game itself can be made the subject of imperfect information. I'd like to note that is actually quite realistic to assume something is known about the cost functionals of the other player. For example, games are commonly used to represent economic choices, where the results have already been recast to deal with money, an artificial medium of exchange. Furthermore, even non-economic games are dealing with people, such that things that are universally true, such as certain needs, can partially determine this structure.

Another common assumption is that the players are rational, and will attempt to do the best for themselves. An interesting violation of this is the player of the multistage game who attempts to bully the other players by forcing the other players into something less than optimal by making high-regret choices if they don't comply. We can say that an extention to these assumptions is that the players play security strategies, in which they try to assure themselves of the maximum expected outcome. Oftentimes, this is all one can do when confronted with single-stage games.

Game theory fits very neatly into the decision theoretic framework we have been setting up. With very few modifications, we can extend the theory in one discipline to cover the entire series. Each of these can be extended to continuous cases as well. Optimal control theory, for example, is more often associated with operations in the continuous world, such that instead of $z_{k+1} = f(x_k, u_k)$, we use $\frac{\delta x}{\delta t} = f(x(t), u(t))$.

Problem Name	Notion of Optimality	Player/DM count	stage count	Nature?	Number of cost functionals
Classical Optimization	optimal	1	1	N	1
Decision Theory	expected optimal	1	1	Y	1
Multiple-Objective Optimality	Pareto optimal	1	1	Y or N	>1
Classical Game Theory Matrix Games	saddle points Nash Equilibrium	>1	1	N	>1
Markov Games	saddle points Nash Equilibrium	>1	1	Y	>1
Optimal Control Theory	optimal policy	1	>1	N	1
Stochastic Control Theory/MDPs	expected optimization	1	>1	Y (prob.)	1
Dynamic/Differential Game Theory		> 1	> 1	Y/N	> 1
Team Theory		>1	>1	Y/N	1

Team theory is totally cooperative, but with multiple information spaces. Given the above, we see that all of these disciplines are special cases of dynamic game theory.

Single Stage, Two-Player, Zero-Sum Games

In zero-sum games, one player is strictly competing against the other player, such that one's loss is the other's gain and the other's gain is the first's loss. Both players act at the same time, such that neither player knows the other player's action before they act. These games can be represented with a single cost matrix, in which one of the players is trying to maximize the matrix, while the other player is trying to minimize the matrix. Non-zero sum games are games in which this is not true, and they need a two matrix representation of cost (unless the game is totally cooperative, of course).

Let us call the two players p_1 and p_2. Let us say that p_1 is the minimizing player, and that p_2 is the maximizing player. Furthermore, let us say that the choice of actions are represented as u^1 and u^2 respectively. We can say that V is the results of a given game. Then, the best that p_1 can do for itself is $\underline{V}$, which is called the lower value, and equals $\max_j \min_i L(u_i^1, u_j^2)$. Similarly, the best that p_2 can do for itself is $\overline{V}$, called the upper value, which equals $\min_i \max_j L(u_i^1, u_j^2)$. For example, consider the zero-sum game:

		P_2, maximizer			
		u_1	u_2	u_3	u_4
	u_1	1	3	3	-2
P_1	u_2	0	-1	2	1
minimizer	u_3	-2	2	0	1

In this game, $\underline{V} = 0$ and $\overline{V} = 2$. We can see that in this game, there is regret for one of the players, who now having seen the move of the other player, wishes to move again. In this game, if p_1 chooses u_2 and p_3 chooses u_3, then the value is 2, and p_1 wishes it had chosen u_3, which has a value of 0, and thus p_1 has a regret of 2. Similarly, we can see that any game will lead to situation such that, should another player know the other player's move ahead of time, could have made a more advantagous move for themselves.

A situation in which there are no regret is when there is a "saddle point." In this case, $V = \underline{V} = \overline{V}$. Such a saddle point is created by the same choice being the best choice for each player, given any choice of the other player. The above game does not have such a saddle point, but the following game does:

		P_2	
	3	5	6
P_1	9	6	3
	1	2	-1
	0	7	4

In that game, $V = 2$. In general, such games with a single saddle point have the following structure:

			$\geq V$			
			$\geq V$			
			$\geq V$			
$\leq V$	$\leq V$	$\leq V$	V	$\leq V$	$\leq V$	$\leq V$
			$\geq V$			
			$\geq V$			
			$\geq V$			

It is also possible for games to have multiple saddle points. Here is a diagram that shows the structure of these games:

			$\geq V$		$\geq V$	
			$\geq V$		$\geq V$	
			$\geq V$		$\geq V$	
$\leq V$	$\leq V$	$\leq V$	V	$\leq V$	V	$\leq V$
			$\geq V$		$\geq V$	
$\leq V$	$\leq V$	$\leq V$	V	$\leq V$	V	$\leq V$
			$\geq V$		$\geq V$	

As aptly pointed out in class, if a saddle point appears in a different row and column than another saddle points, that necessitates more saddle points. Observe the above diagram. We see that values at the saddle point must be both $\geq V$ and $\leq V$, and therefore must be V. Therefore, there is no chance of mismove, and there will never be regret in a full-information, zero-sum game with at least one saddle point.

Unfortunately, we have observed that not all single stage games have saddle points, so regret will exist. But what about multiple-stage games? It turns out there is a randomized, or mixed, strategy which will yield an expected value for the game, such that over the long run there is no regret. This result will be shown after we explain how mixed strategies fit into our terminology and how we represent them mathematically.

We can understand all zero-sum games in this terminology. In single-stage games with two ordinary players, we say that both players will use a pure strategy. We can think of a single stage game against nature as the normal player moving with a pure strategy and nature as moving with a mixed strategy. Then, in multiple stages games, both players use mixed strategies to their benefit.

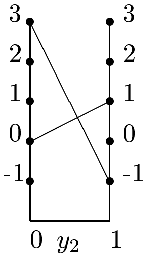
A strategy is represented by probability vectors, where each probability in the vector corresponds to an action available to the player, and the vector has the cardinality of the number of actions. Thus, we say that P_1 has a vector $y = [y_1 y_2 y_3 \ldots y_n]^T$, where $y_i \geq 0$ and $\Sigma z_i = 1$, and that P_2 has a similar z vector. Then, let A be the loss matrix. Then, the expected loss for P_1 is $E[L_1] = y_T A z = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} y_i z_j$. Since these games are zero-sum, we can say that the expected loss for P_2 is $E[L_2] = -E[L_1]$.

Choosing the mixed secure strategy gives an average upper value, $\overline{V}_m = \max_z y^{*T}Az \leq \max_z y^T Az \forall y$ where y^* is the best possible pick for P_1. z^* for P_2 is similarly the best pick for P_2, where instead we minimize over y. It turns out that $\underline{V} \leq \underline{V}_m \leq \overline{V}_m \leq \overline{V}$

Well, it's all well and good to say that there is always a mixed saddle point, but how does one find it? It turns out that we can perform linear programming on the probability space, which is similar to the information space. To facilitate this, we note that $y^T Az$, with either Az or yA (depending on the player) held constant, is a vector of linear equations. So let us consider a game where each player has two actions. Note that this game has no pure saddle point.

		P_2	
		u_1	u_2
P_1	u_1	3	0
	u_2	-1	1

We can then draw the probability space for P_1 in terms of its y vector, which we can express with a single entry for the probability of choosing u_2, or y_2, knowing that $y_1 = 1 - y_2$. We draw one line for each pure choice of the other player.



Here, we observe that $y_2^* = \frac{3}{5}$ and $y_1^* = \frac{2}{5}$. $V_m = \frac{3}{5}$, which has nothing to do the probabilities of y, but is instead where the lines cross, and more importantly is the location of the minimum point. The linear programming techniques used to exploit this is too big a subject to be covered here, so just look on Google like everybody else.

In this two by two case, we find that $z_1^* = \frac{1}{5}$ and $z_2^* = \frac{4}{5}$. It is necessary to draw a separate diagram to discover this. It seems counterintuitive, since this player is moving more often in favor of the lesser reward, and at a greater frequency than the other player! But, observe that the penalty is also greater. If a player has more than two actions, then we will be drawing planes or hyperplanes instead of lines, where corners in the space correspond to pure strategies (in P_1's case this means that some $y_i = 1$, and that there is a plane for the length of z).

But, how does this work no matter what? Suppose the worst case, that the other player reads in the paper what your strategy is, and how your probability generator works. Then, you are clearly in trouble, because the player then is told of what your move will be and can react accordingly. But, given that they actually don't find out how the probabilities work, or that they are truly random. Then, they cannot take advantage of such a strategy. We can observe this directly by observing that the only way they could take advantage would be to switch to a different strategy other than their optimal mixed strategy, but then they would no longer be at the optimal point. This isn't to say that there aren't better exploitations of a non-optimal strategy, however!

Non-Cooperative, Non-Zero Sum, Two-Player Game Theory

Again, suppose a single stage, 2 player game. Now, however, the two players have independent loss functions, which we can respresent with two matrices. Let A correspond to P_1's loss matrix and B correspond to that of P_2, where both players are trying to find the minimum. One player chooses the row, usually P_1 and the other, P_2, chooses the column. The now-infamous Nash equilibrium is the result where both players have no regret. The value at row i^*, column j^* is a Nash equilibrium if $a_{i^*j^*} \leq a_{ij^*} \forall i \in 1, \ldots, n$ and $b_{i^*j^*} \leq b_{ij^*} \forall i \in 1, \ldots, n$. It's like a saddle point where the saddle is shared across two boards. So if we have an A and B of the form:

$\leq V$ $\leq V$ $\leq V$ V $\leq V$ $\leq V$ $\leq V$

	$\leq V$	
	$\leq V$	
	V	
	$\leq V$	
	$\leq V$	

So, given that each player played what they did, there is no regret, in the Nash equilibrium. Suppose the matrices:

A, the row picker	
2	3
1	0

B, the column picker	
1	0
2	-1

We can see that there are two Nash equilibrium, one at the top left, and one at the bottom right. But, if the players go for different Nash equillibrium, then there is trouble! Both players have regret, having been able to do a point better having better. Even though both are Nash equilibria, you can do a little better collectively. -1 and 0 is better than 1 and 2, since both players do better. So, we can say that between two Nash equilibria with coordinates (i_1, j_1) and (i_2, j_2), 1 is better than 2 if $a_{i_1j_1} \leq a_{i_2j_2}$ and $b_{i_1j_1} \leq b_{i_2j_2}$ and at least one of the inequalities is strict. A Nash equilibrium is called admissible if there are no better Nash equilibrium. We can say that such an equilibrium point is Pareto optimal, as well, given our previous definitions. This is definitely a criterion for attempting to eliminate equilibrium points for regret avoidance. Unfortunately, some games have multiple admissible equilibria. For example, consider the following game.

-2	1
-1	-1

-1	1
2	-2

Both (-1,-2) and (-2,-1) are results of this game, and thus regret will occur unless some sort of communication and negotiation are possible. Even then, if the game only occurs one time, there may be very little to bargain about. Instead, the players may choose to play security strategies. In this P_1 selects the row in A with the lowest max, while P_2 selects the column in B with the lowest max.

Another phenomenon that occurs is for the actions to be interchangable. This happens when the game is actually a zero-sum game (B = -A).

One of the most classic illustrations of Nash equilibrium resulting in a non-optimal result, and of security strategies being chosen, is the "Prisoners' Dilemma." In this scenario, two bank robbers are about to get caught, so they get rid of the loot, the guns, the plastic masks of Ronald Reagan and George Bush, and start loitering around playing cards. When the police come, they say "Yeah, we saw two armed ex-presidents going that way," pointing vaguely off into the distance. The police arrest them anyhow and find, while searching them, besides expired credit cards and chewing gum and moustache wax and toenail clippers and so forth, bootleg Microsoft software. So, they bring them in, put them in different rooms with concrete walls and lit only by a lamp in the center that can be brandished dramatically. So, here's the deal: if they both confess, they'll get the recommended minimum for bank robbery, 8 years. If they both claim to be innocent, then they will be changed for sofware piracy and given a 2 year sentence. If one of them confesses, however, that one will be let loose for turning evidence while the other will get the maximum, 20 years. So, the matrices for the game look like:

P_1's matrix

	P_1 claims innocent	P_1 confesses
P_2 claims innocent	2	0
P_2 confesses	8	20

P_2's matrix

	P_1 claims innocent	P_1 confesses
P_2 claims innocent	2	8
P_2 confesses	0	20

We can see that no matter what the other player picks, the best move for each player is to confess, which will move them from 2 to 0 or from 20 to 8.

Another example of a "Prisoner's Dilemma" non-cooperative negative result is advertising. If two strong competitors (say Coca-Cola and Pepsi) did a minimal amount of advertizing, they would both do well, maybe not selling as much, but they are overall better off given that they don't have to pay the expense of advertising. But if one has a big ad campaign, and the other doesn't, then the company that didn't will get killed. So, both must advertise.

2-player, Non-zero Sum, Multi-stage Games

The Nash equilibrium also always has at least 1-mixed equilibrium, which shouldn't surprize anyone, since all it means is that there is an expected pure space minimum. We say that y^*, z^* is a Nash equlibria if $y^{*T}Az^* \leq y^TAz, \forall y \in Y$ and $y^{*T}Bz^* \leq y^{*T}Bz, \forall z \in Z$. The results might be better, worse, or incomparable to pure Nash equilibrium. One can use quadratic programming to compute such mixed equlibrium.

Variations and Terminology

Even outside of these traditional variations, there are many interesting extensions. For example, suppose some players are allowed to move ahead of other players based on the state of the game. A simple variation on this is the Stackelberg game, in which one of the players always moves first. The first player, or leader, can force the best row for it on the second player, or follower. We can see that the follower only has one reasonable choice, which the first player can be assured of. Another interpretation of the Stackelberg idea is that although the first player does not move first, it announces a strategy to player two, who will then move optimally within the bounds of the agreement, which supposedly has been chosen for mutual benefit. Pure Stackelberg equillibria always exist for normal form games.

One can say that a strategy is open loop or closed loop. An open loop strategy considers only the first state ($\eta_k^i = \{x_1\}$), while a closed loop includes all of the previous states ($\eta_k^i = \{x_1, \ldots, x_k\}$). We can also represent a closed-loop with imperfect state ($\eta_k^i = \{y_1^i, \ldots, y_k^i\}$). One might also know of a memoryless perfect state ($\eta_k^i = \{x_1, x_k\}$) in which we only know the most recent and first states. We may have a feedback strategy ($\eta_k^i = \{x_k\}$) using only the most recent perfect information. We can also have an imperfect feedback state ($\eta_k^i = \{y_k^i\}$). We can also talk about delayed memory, where the one step case is ($\eta_k^i = \{x_1, \ldots, x_{k-1}\}$). And, we can say all kinds of combinations of these where we remember a finite history, where we remember aggregates of states, where we can see the current state but our action won't come into effect until later, and any number of other things.