CSC304: Algorithmic Game Theory and Mechanism Design
Fall 2016

Allan Borodin
Tyrone Strangway and Young Wu (TAs)

September 21, 2016

Lecture 5

- Announcements
  - Tutorial this Friday in SS 1086
  - First 3 questions for assignment 1 have been posted.
  - Another talk of possible interest: Thursday, September 29, there will be a seminar "When Should an Expert Make a Prediction?" by Amir Ban. The talk will take place in the first floor conference room of the Fields Institute, 222 College Street. The time will either be at noon or at 1PM.
- Today's agenda
  - Expressing the value of a zero-sum game as an LP
  - The LP duality theorem;
  - Yao's minimax theorem;
  - The hide and seek game; maxmimum matching in a biparite graph
  - Briefly returning to general sum games; many player games

Two important computational applications of the minimax theorem

We ended the last lecture mentioning two important applications of the minimax theorem.

1. The first application is actually an equivalent result, namely the (LP) **duality theorem** of linear programming (LP). Linear programming is one of the most important concepts in combinatorial optimization (leading to efficient optimal and approximation algorithms) and LP duality is arguably the central theorem of linear programmming. Since LPs can be solved optimally in polynomial time, this will imply that (unlike general-sum NEs), we can always solve zero-sum games (i.e. find the mixed strategies that yield the value of the game).
2. The second application is The Yao Principle which is a direct consequence of the minimax theorem and is a basic tool in proving "negative" results for randomized algorithms.

Integer programming, linear programming and LP duality

An integer program (IP) (resp. LP) formulates an optimization problem as the maximization or minimization of a linear function of integral (resp. real) variables subject to a set of linear inequalities and equalities.

Most combinatorial optimization have reasonably natural IP (and sometimes LP) formulations. But solving an IP is an NP complete problem so that one does not *in general* expect efficient algorithms for optimally solving IPs. However, many IPs are solved "in practice" by IPs and there are classes of IPs that do have worst case efficient algorithms.

Another important idea in approximately solving IPs is to relax the integral constraints to real valued constraints (e.g. $x_i \in \{0, 1\}$ is relaxed to $x_i \in [0, 1]$) and then "rounding" the fractional solution to an integral solution.
Note: if the objective function and all constraints have rational coefficients, then we can assume rational solutions for an LP.

Interger programming, linear programming and LP duality continued

In contrast to IPs, LPs can be optimally solved in polynomial time (e.g. by interior point methods) . In addition to having polynomial time methods for solving LPs, Dantzig's simplex method (with different pivoting rules) is widely used "in practice" and often provides efficient solutions.

Furthermore, one can often use LP duality theory to yield good approximations (and provable limitations for a given LP formulation) without solving the LP.

We next show how optimal mixed strategies can be solved by LPs, and provide a precise statement of LP standard form and LP duality.

LP formulation of optimal mixed strategies

Consider a two person zero-sum game represented by an $m \times n$ matrix A. To verify that the row player's mixed strategy $\mathbf{x}$ has an expected gain of at least value v, it is sufficient (and necessary) to show that the expected gain is at least v for every pure strategy of the column player. That is, in vector notation, we need to guarantee $A^T\mathbf{x} \geq v\mathbf{e}$ where $\mathbf{e}$ is the all 1's vector.

The goal of the row player can then be stated as the following LP:

Maximize v Subject to: $A^T\mathbf{x} \geq v\mathbf{e}$

$\sum_{1 \leq i \leq m} x_i = 1$

$x_i \geq 0$ for all i; $1 \leq i \leq m$

Similarly, the optimal mixed strategy for the column player is to:

Minimize v Subject to: $A\mathbf{y} \leq v\mathbf{e}$

$\sum_{1 \leq j \leq n} y_j = 1$

$y_j \geq 0$ for all j: $1 \leq j \leq n$

An example of how to compute value (and mixed strategies) for zero sum game

Recall the Penalty Kick Game that we saw in Lecture 4.

kicker \ goalie	L	R
L	0.58	0.95
R	0.93	0.70

Figure: Probability matrix for scoring based on data from professional games

The LP formulation for this game is as follows:

Maximize v Subject to: $.58x_1 + .93x_2 \geq v$

$\sum_{1 \leq i \leq 2} x_i = 1$

$x_i \geq 0$ for all i; $1 \leq i \leq 2$

Using the LP solver

Using the LP solver for the penalty kick data matrix, we have :

max: v;
.58 x1 + .93 x2 >= v;
.95 x1 + .70 x2 >= v;
x1 + x2 = 1;

Results of the solver:
Value of objective function: 0.795833
Actual values of the variables:

v	0.795833
x1	0.383333
x2	0.616667

LP standard form and LP duality

The minimax theorem tells us that these two LPs, must have the same optimal value v. This is an example (after some messaging to place the constraints iin standard form) of LP duality.

LP maximization in standard form

Any maximization LP can be stated in the following standard form:

Maximize $\mathbf{c}^T\mathbf{x}$ Subject to $A\mathbf{x} \leq \mathbf{b}$

$\mathbf{x} \geq \mathbf{0}$

We will refer to this LP as the *primal* LP.

The minimization *dual* LP (in standard form) of the above primal

Minimize $\mathbf{b}^T\mathbf{y}$ Subject to $A^T\mathbf{y} \geq \mathbf{c}$

$\mathbf{y} \geq \mathbf{0}$

We could alternatively start with a minimization problem in standard form as the primal and then the dual is a maximization problem in standard form. Moreover, the dual of the dual is the primal.

LP duality theorem of linear programming

Let P (resp. D) denote the primal and dual LPs (in standard form) and let $\mathcal{F}(P)$ (resp $\mathcal{F}(D)$) be the feasible sets for P (resp. D).

The duality theorem

$\mathcal{F}(P)$ has a finite optimum iff $\mathcal{F}(D)$ has a finite optimum. Furthermore, if $\mathbf{x}^*$ and $\mathbf{y}^*$ are optimal solutions for the primal and dual LPs, then $\mathbf{c}^T\mathbf{x}^* = \mathbf{b}^T\mathbf{y}^*$.

Similar to the easy direction of the minimax theorem, there is a "weak duality" direction. Assume again that $\mathcal{F}(P)$ (and therefore resp. $\mathcal{F}(D)$) has a finite optimum $\mathbf{x}^*$ (resp $\mathbf{y}^*$). Then

The weak duality theorem

$\mathbf{c}^T\mathbf{x}^* \leq \mathbf{b}^T\mathbf{y}^*$.

The weak direction is often used to provide an upper bound (resp. a lower bound) on the optimal value of a maximization (resp. minimization) problem.

Motivating duality

Example in V. Vazirani's "Approximation Algorithms" text:

Primal

minimize $7x_1 + x_2 + 5x_3$

subject to

- (1) $x_1 - x_2 + 3x_3 \geq 10$
- (2) $5x_1 + 2x_2 - x_3 \geq 6$
- $x_1, x_2, x_3 \geq 0$

Dual

maximize $10y_1 + 6y_2$

subject to

$$y_1 + 5y_2 \leq 7$$
$$-y_1 + 2y_2 \leq 1$$
$$3y_1 - y_2 \leq 5$$
$$y_1, y_2 \geq 0$$

Adding (1) and (2) and comparing the coefficient for each x_i, we have:
$7x_1 + x_2 + 5x_3 \geq (x_1 - x_2 + 3x_3) + (5x_1 + 2x_2 - x_3) \geq 10 + 6 = 16$
Better yet, $7x_1 + x_2 + 5x_3 \geq 2(x_1 - x_2 + 3x_3) + (5x_1 + 2x_2 - x_3) \geq 26$
For an upper bound, setting $(x_1, x_2, x_3) = (7/4, 0, 11/4)$
$7x_1 + x_2 + 5x_3 = 7 \cdot (7/4) + 1 \cdot 0 + 5 \cdot (11/4) = 26$
This proves that the optimal value for the primal is at least 26.
Setting $(y_1, y_2) = (2, 1)$, the dual is at most 26 so optimal value $= 26$.

Randomized algorithms and Yao's minimax principal

We will soon state Yao's principle which provides a basic methodology for analyzing the "cost" of randomized algorithms.

Suppose we are intetrested in considering a finite class $\mathcal{A}$ of deterministic or randomized algorithms for a computational problem restricted to a finite set of inputs. More specifically, lets restrict attention to inputs of "size" n.

For example, the *size* of the input can refer to the number of bits in a binary representation of the input.

The *cost* of a deterministic algorithm can be the number of time steps, or the number of memory cells being used. If $\mathcal{A}$ is solving (say) a minimization problem, the cost could be the largest (i.e. worst) value produced by the algorithm (when restricted) to inputs of size n).

We can think of a randomized algorithm as a distribution $\mathcal{R}$ over deterministic algorithms and the cost of such an algorithm as the expected cost where the expectation is over the randomness used.

Randomized complexity and distributional complexity

One measure of interest is the *randomized complexity* of the given problem; namely, the worst case (over all inputs of size n) of the cost of the best randomized algorithm (from the class $\mathcal{A}$). That is,

$$\min_{\mathcal{R}} \max_{w:size(w)=n} \mathbb{E}_{A\in\mathcal{R}}[cost(A, w)]$$

Another measure of interest is the worst case *distributional complexity* of the given problem; namely, the worst case (over all distributions of inputs of size n) of the cost of the best deterministic algorithm. That is,

$$\max_{\mathcal{D}} \min_{A\in\mathcal{A}} \mathbb{E}_{w\in\mathcal{D}}[cost(A, w)]$$

Yao's minimax principal

We can now think of a zero-sum game between a row player whose strategy is to choose an input from a distribution $\mathcal{D}$ and a column player whose strategy is to choose an algorithm from a distribution $\mathcal{R}$ of algorithms. Yao's principle (from a seminal 1977 paper by Andy Yao) is then the application of the minimax theorem to this game. Namely,

Yao's principle

$$\max_{\mathcal{D}} \min_{A \in \mathcal{A}} \mathbb{E}_{w \in \mathcal{D}}[cost(A, w)] = \min_{\mathcal{R}} \max_{w:size(w)=n} \mathbb{E}_{A \in \mathcal{R}}[cost(A, w)]$$

In words, the distributional complexity is equal to the randomized complexity. It is usually the "weak direction" of Yao's principle that is used to establish lower bounds on the randomized complexity; that is, to show a lower bound $C(n)$ on the randomized complexity of a problem it suffices to establish a distribution $\mathcal{D}$ on inputs such that every deterministic algorithm (in the class $\mathcal{A}$) will have expected cost at least $C(n)$.

Zero-sum games on graphs

Chapter 3 provides a nice opportunity to introduce the use of graphs (and graph theory) in many aspects of game theory, mechanism design and social choice. One particular class of graphs, *bipartite graphs*, are often of interest.

For example, later we will discuss *unit demand auctions* for a set of distinct items, which can be modeled as an edge weighted bipartite graph. We will also be considering "stable matchings" which are used (for example) to match residents to hospitals and also is refered to stable marriages. For now we will just mention one example (the "hide and seek game") and its relation to maximum matchings in bipartite graphs, a topic discussed in section 3.2 of the KP text.

Who is and who is not familiar with basic graph theoretic concepts?

Zero-sum games on graphs

Chapter 3 provides a nice opportunity to introduce the use of graphs (and graph theory) in many aspects of game theory, mechanism design and social choice. One particular class of graphs, *bipartite graphs*, are often of interest.

For example, later we will discuss *unit demand auctions* for a set of distinct items, which can be modeled as an edge weighted bipartite graph. We will also be considering "stable matchings" which are used (for example) to match residents to hospitals and also is refered to stable marriages.
For now we will just mention one example (the "hide and seek game") and its relation to maximum matchings in bipartite graphs, a topic discussed in section 3.2 of the KP text.

Who is and who is not familiar with basic graph theoretic concepts?

I am assuming everyone is familiar with basic graph theoretic concepts but if not I will provide a quick tutorial (now or in a tutorial).

The hide and seek game; section 3.2 of KP

The hide and seek narrative (slightly sensationalized): example 3.2.1 in KP text)

A terrorist is hiding in one of a few *known* safe houses in a city (basically) organized by east-west streets and north-south avenues (like Manhatten in NYC, Seattle, etc.) The safe houses are located at the intersections. An FBI agent plans to drive down one of the streets or avenues and will be able to identify the location of the terrorist (say with some probability) and has an expected winning payoff (which we set to 1) if she drives down the street or avenue where the terrorist is hiding; otherwise, the payoff is 0.

- What strategy should the FBI agent take?
- What strategy should the terrorist use?

The graph theoretic model and optimal strategies

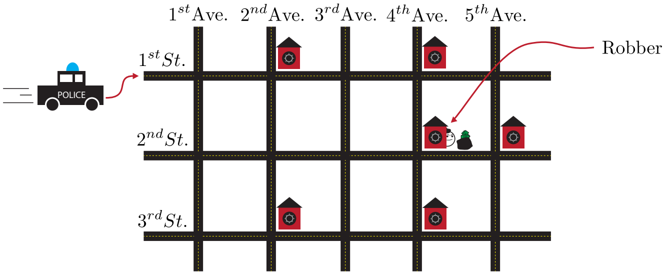

Figure 3.5. The figure shows an example scenario for the *Hide-and-Seek* game. In this example, the robber chooses to hide at the safehouse at the intersection of 2nd St. and 4th Ave., and the cop chooses to travel along 1st St. Thus, the payoff to the cop is 0.

One can represent the problem using a bipartite graph $G = (U \cup V, E)$ where U represents the set of streets and V the set of avenues and then the set of edges $E \subseteq U \times V$ corresponds to the intersections of the safe houses.

The strategies for the FBI and the terrorist

- Since the FBI wants the best chance to find the safe house by traversing a single street or row, its best strategy is to first find a minimum size *line cover* C; that is, a set of streets and avenues (i.e. $C \subseteq U \cup V$) that covers all possible safe hiding places (i.e. that is, C "covers" all the edges in G. In graph theory terms, the FBI is seeking a *minimum vertex cover* for the graph G. Then the strategy would be to uniformly at random choose a "line" in C.
- The strategy of the terrorist is to first find a maximum size set M of safe houses such that none lie on the same street or avenue and then choose uniformly at random from M. In graph terms, M is called a maximum matching.

What does graph theory tell us?

König's Theorem (which uses Hall's Marriage Theorem) shows that in a biparite graph, the size of a minimum vertex cover is equal to the size of a maximum matching.

Aside: It is known how to efficiently compute a maximum matching in any graph and this can be used to find a minimum size vertex cover in a bipartite graph. On the other hand, computing a minimum size vertex cover for an arbitrary graph is an "NP-complete problem" (although a problem that is often solved well "in practice").

Hall's Marriage Theorem

Let $G = (U \cup V, E)$ be a bipartite graph and (without loss of generality) let $|U| = m \leq n = |V|$. Then a necessary and sufficient condition for having a maximum matching of size m is that the neighborhood $Nbhd(S) = \{v \in V | \exists (u, v) \in E\}$ of every subset $S \subseteq U$ has size at least $|S|$.

Alternative interpretations for Konig's Theorem

- If we formulate a natural $\{0,1\}$ integer program IP (i.e. all coefficients are in $\{0, 1\}$) for maximum matching in a bipartite graph, and P is the LP relaxation of this IP, then it is known that the IP optimum is equal to the LP optimum.

  Note solving IPs is an NP hard optimzation probelm and usually an LP relaxtion has a much better optimum

Alternative interpretations for Konig's Theorem

- If we formulate a natural $\{0,1\}$ integer program IP (i.e. all coefficients are in $\{0,1\}$) for maximum matching in a bipartite graph, and P is the LP relaxation of this IP, then it is known that the IP optimum is equal to the LP optimum.

  Note solving IPs is an NP hard optimzation probelm and usually an LP relaxtion has a much better optimum

  The dual of this primal P is a formulation of the vertex cover problem.
- We can also express the bipartite graph as a vertex-edge adjacency $\{0,1\}$ matrix. This then directtly expresses the hide and seek game as a zero-sum game where the row player is the vertex (i.e. FBI player) trying to maximize the value of a row strategy which is the number of 1's in that row. The column player (the terrorist) is trying to minimize the chance of being found.

Returning to general-sum games

While we mainly discussed two person general-sum games at the start of the course, we did indicate that the concepts being introduced applied to games with a finite but arbitrary number of players. We will now state those concepts in the context of $k \geq 2$ players.

Let S_i be the set of possible strategies for the i^{th} player. A pure strategy profile $(s_1, s_2, \ldots, s_k)$ specifies the pure strategies of the k players with $s_i \in S_i$ being a strategy of the i^{th} player. We follow standard notation and let $\mathbf{s}_{-i}$ denote the pure strategies for all players except player i; that is, $\mathbf{s}_{-i} = (s_1, \ldots s_{i-1}, s_{i+1}, \ldots, s_k)$ and then let $(s_i, s_{-i}) = (s_1, s_2, \ldots, s_i \ldots, s_k)$.

More generally, a mixed strategy profile $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ specifies the mixed strategies of the k players with $\mathbf{x}_i$ being the strategy of the i^{th} player. And similarly, $\mathbf{x}_{-i}$ denotes the mixed strategies all players except player i so that $(\mathbf{x}_i, \mathbf{x}_{-i}) = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_k)$. We let $\mathbf{x}_i(s)$ denote the probability that the i^{th} player chooses strategy $s \in S_i$.

Pure and mixed NE and recalling Nash's theorem

$u_j(\mathbf{s})$ denotes the payoff (or utility) of agent j when the players are playing the pure strategy profile $\mathbf{s}$. A pure strategy profile $(s_1^*, \ldots, s_k^*)$ is a pure NE if for each player j and for each strategy s_j for player j , we have:

$$u_j(s_j^*,\ s_{-j}^*) \geq u_j(s_j, \mathbf{s}_{-j}^*)$$

For a mixed strategy profile $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$, the utility for player j is

$$u_j(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) = \sum_{s_1 \in S_1, \ldots s_k \in S_k} \mathbf{x}_1(s_1) \cdots \mathbf{x}_k(s_k) u_j(s_1, \ldots, s_k);$$

that is, $u_j(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$ is the expected value of the j^{th} player when the players are using the mixed strategy profile.
Finally, we can define a mixed NE for an arbitrary number $k \geq 2$ of players. Namely, $(\mathbf{x}_1^*, \ldots \mathbf{x}_k^*)$ is a mixed NE if

$$u_j(\mathbf{x}_j^*, \mathbf{x}_{-j}^*) \geq u_j(\mathbf{x}_j, \mathbf{x}_{-j}^*)$$

for all j and all mixed strategies $\mathbf{x}_j$ for the j^{th} player.

How to visualize and analyze normal form games for many players

While in principle, we can just specify all the relevant probability vectors and the utitilies for each pure strategy profile. Nash's theorem applies to any finite number of players each having finitely many strategies. So there must be at least one mixed (possibly pure) NE. And we can again use the principle of indifference to find possible mixed NE.

How to visualize and analyze normal form games for many players

While in principle, we can just specify all the relevant probability vectors and the utitilies for each pure strategy profile. Nash's theorem applies to any finite number of players each having finitely many strategies. So there must be at least one mixed (possibly pure) NE. And we can again use the principle of indifference to find possible mixed NE.

But in practice, this is not so easy to manage. For 3 player games there is a standard way that one often uses to display such games. Namely for say the 3rd player, we proved a matrix for each possible pure strategy of the 3rd player where each entry of these matrices consists of a vector of the payoffs fpor the 3 players.

This is how the text specifies the *Pollution game* in section 4.3 amongst 3 players where each player has two strategies, namely to purify or to pollute. The analysis whill the lead to some quadratic equations form which the possible NE are calculated. It then utilizes the principle of indifference