Lecture 2: Minimax theorem, Impagliazzo Hard Core Lemma

Topics in Pseudorandomness and Complexity Theory (Spring 2017)
Rutgers University
Swastik Kopparty
Scribe: Cole Franks

Zero-sum games are two player games played on a matrix $M \in \mathrm{Mat}_{m\times n}(\mathbb{R})$. The row player, denoted R, chooses a row $i \in [m]$ and the column player C chooses a column $j \in [n]$, simultaneously. The payoff to the row player is M_{ij} and the payoff to the column player is $-M_{ij}$ (hence the game is "zero-sum"). We can also consider randomized strategies, where R chooses a probability distribution p on $[m]$, which can be written as a vector in $\mathbb{R}^m$, and the column player chooses a probability distribution q on $\mathbb{R}^n$. The expected payoff to the row player, denoted $E(p,q)$, is thus equal to

$$p^T M q = \sum_{ij} M_{ij} p_i q_j.$$

Theorem 1 (Von Neumann Minimax Theorem)**.** *For a linear game there is a value V such that*

1. *There exists p^* such that for all q, $E(p^*, q) \geq V$ and*
2. *there exists a q^* such that $E(p, q^*) \leq V$ for all p.*

In words: R can guarantee that the expected payoff to R is at least V, while C can guarantee that the expected payoff to R is at most V. This V is called the value of the game.

Proof. We say that R can guarantee that the expected payoff to R is at least α if there exists a probability distribution $p \in \mathbb{R}^m$ such that for all probability distributions $q \in \mathbb{R}^n$, $E(p,q) \geq \alpha$.

Lemma 2. *R can guarantee that the expected payoff to R is at least α if and only if there is a p such that for all $j \in [m]$ $E(p, e_j) \geq \alpha$.*

Proof. The "only" if direction is clear, because the deterministic strategies is a subset of the randomized strategies. To prove the "if" direction, suppose we have such a p. Note that for any q,

$$E(p,q) = \sum_{i\in[m]} \sum_{j\in[n]} p_i M_{ij} q_j = \sum_{j\in[n]} q_j E(p, e_j) \geq \alpha,$$

because q is a probability distribution on $[n]$. □

Lemma 3. *Given X, Y closed, disjoint convex sets in $\mathbb{R}^n$, there is a hyperplane that separates X and Y, that is, a unit vector u and a number t such that either for all $x \in X$ and $y \in Y$, $\langle u, x\rangle \leq t$ and $\langle u, y\rangle > t$ or for all $x \in X$ and $y \in Y$, $\langle u, x\rangle < t$ and $\langle u, y\rangle \geq t$.*

Proof sketch. Suppose X and Y are compact. Consider $\inf_{x\in X, y\in Y} d(x,y)$; we know this distance is achieved by some x_0 and y_0 because it is a continuous function on $X \times Y$ which is also compact. One can check that the hyperplane bisecting the segment xy separates X and Y. Suppose X and Y are only closed; then $X_N \equiv X \cap \{x : |x_i| \leq N\}, Y_N \equiv \cap\{x : |x_i| \leq N\}$ are compact sets for each natural number N. There is a separating hyperplane H_N for each X_N and Y_N; H_N can be described by a unit vector u_N and a threshold t_N such that $H_N = \{x : \langle x, u_N\rangle = t_N\}$. Suppose we always pick u_N so that X_N is on the negative side of H_N. t_N is in fact the distance from the origin of H_N, which must be bounded above because some points $x \in X$ and $y \in Y$ are at a finite distance from the origin and hence cannot be separated by any hyperplane that is arbitrarily far from the origin despite being contained in X_N and Y_N for N sufficiently large. Hence $\{(u_N, t_N)\}_{N\in\mathbb{N}}$ resides in a compact subset of $\mathbb{R}^{n+1}$, so it has a convergent subsequence. Let (u,t) be the limit of this subsequence. For each $x \in X$, and for N sufficiently large, $\langle x, u_N\rangle < t$, and $\langle x, u\rangle = \lim_{N\to\infty}\langle x, u_N\rangle \leq t$. Similarly for y, we have $\langle u, y\rangle \geq t$. The intersections $X \cap H$ and $Y \cap H$ are closed convex sets. If either is empty, we are done. If neither is empty, then by induction there is a hyperplane $G \subset H$ that separates $X \cap H$ and $Y \cap H$ as in the theorem. Without loss of generality suppose G does not intersect $X \cap H$. By translating, assume G (and so H) is through the origin with normal v, so that $\langle x, v\rangle < 0$ for $x \in X$ and $\langle y, v\rangle \geq 0$ for $y \in Y$. Let $u' = u + .1v$. Now $\langle x, u'\rangle = \langle x, u + .1v\rangle < 0$ and similarly $\langle y, u'\rangle \geq 0$. Apply the inverse of the translation to obtain a new hyperplane (u', t') that separates X and Y in the way required by the theorem. □

Take some value α such that R cannot guarantee expected value at least α. Then for all p there exists j such that

$$\sum_i p_i M_{ij} \leq \alpha.$$

Let $X = \{u \in \mathbb{R}^n : u_j \geq \alpha \forall j \in [n]\}$ and $Y = \{pM : p \text{ is a probability distribution on } [m]\}$. R cannot guarantee α if and only if X and Y are disjoint. One needs to check that these are convex and closed. We claim that if R cannot guarantee α then there is a hyperplane separatng X and Y. That is, there is $l \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that for all $x \in X$, $\langle l, x\rangle > b$ and for all $y \in T$, $\langle l, y\rangle < b$. Observe the following:

1. Every entry of l is nonnegative. Otherwise you could make the corresponding entry in $x \in X$ arbitrarily large while leaving all others equal to α. This would violate $\langle l, x\rangle > b$.
2. Replace l from the separating hyperplane by $q = \frac{l}{\sum l_i}$ so that q is a probability distribution, and replace α by $b = \frac{\alpha}{\sum l_i}$. So now for all $x \in X$, $\langle q, x\rangle > b$ and for all $y \in T$, $\langle q, y\rangle < b$.
3. $b \leq \alpha$, which one can see by applying $\langle q, x\rangle > b$ for $x = \alpha\vec{1}$.

Items 2 and 3 show that for all $p \in \mathbb{R}^m$, $\langle q, pM\rangle < \alpha$, or $E(p,q) < \alpha$. In other words, if C plays q she guarantees that the expected payoff for R is $< \alpha$.

We know that those values R can guarantee comprise a left half interval, and those values C can guarantee comprise a right half interval. What we've shown is that for any α outside R's left half interval is in the interior of $C's$ right interval. Because our proof was symmetric in R and C, any α outside $C's$ right half interval is in the interior of R's left half interval. Together these imply the

intervals are closed and intersect in exactly one point.

□

Theorem 4 (Impagliazzo Hard Core Lemma)**.** *For all $\epsilon, \delta > 0$, and $\Omega(n/\epsilon^2 \log(n/\epsilon^2)) = s \leq 2^{.9n}$, suppose $f : \{0,1\}^n \to \{0,1\}$ such that for all circuits C with size s*

$$\Pr_{x \in \{0,1\}^n}[C(x) = f(x)] < 1 - \delta.$$

Then there exists a "hard core" $H \subset \{0,1\}^n$, $|H| \geq (2\delta)2^n - O(\sqrt{\delta 2^n})$ such that for all circuits C with size at most $\epsilon^2 \delta s/5n$,

$$\Pr_{x \in H}[C(x) = f(x)] < 1/2 + \epsilon.$$

Proof. Consider a linear game played by two players S (sets, row) and C (circuit, column). S chooses from sets of $\{0,1\}^n$ of size *exactly* $(2\delta)2^n$, which we will call "Large sets" and C chooses from circuits of size at most s', which we will call "Small circuits". The matrix M of payoffs to the circuit player has entries $M_{S,C} = \Pr_{x \in S}[C(x) = f(x)]$.

By the Von Neumann Minimax Theorem and Lemma 2, we are in one of two cases: Either

1. There is a distribution μ_s on Large sets such that for all Small circuits C of size at most s',

$$\mathbb{E}_{S \sim \mu_s}[M_{S,C}] \leq 1/2 + \epsilon,$$

   or

2. There is a distribution μ_c on Small circuits such that for all Large sets

$$\mathbb{E}_{C \sim \mu_c}[M_{S,C}] \geq 1/2 + \epsilon.$$

0.1 Case 1:

We want to obtain a *single* Large set from the distribution μ_s on which no circuit agrees well with f. Define

$$\nu(x) = \Pr_{S \in \mu_s}[x \in S]$$

Then $\sum_{x \in \{0,1\}^n} \nu(x) = (2\delta)2^n$, the expected size of S, by linearity of expectation.
To choose H, for each $x \in \{0,1\}^n$, add x to H independently with probability $\nu(x)$. Then

1. $Pr[|H| \geq (2\delta)2^n(1-\eta)] \geq 1 - e^{-\eta^2(2\delta)2^n/3}$. (The sum of independent Bernoulli's is controlled by the multiplicative Chernoff bound; see Alon and Spencer Corollary A.1.14 or Wikipedia). We will end up choosing η so that this probability is close to one.

2. With high probability over choice of H,

$$\Pr_{x \in H}[f(x) = C(x)] \leq 1/2 + \epsilon/2$$

To prove 2, fix a circuit C. Let Y be the random variable that is the number of agreements between f and C on H. Then

$$\mathbb{E}_H Y = \sum_{x\in\{0,1\}^n} \Pr[x \in H] 1_{f(x)=C(x)} = \sum_{x\in\{0,1\}^n} \nu(x) 1_{f(x)=C(x)} \tag{1}$$

again by linearity of expectation. $M_{S,C} = \frac{1}{(2\delta)2^n} \sum_{x\in S} 1_{f(x)=C(x)}$, so

$$\mathbb{E}_{S\sim\mu_s}[M_{S,C}] = \sum_S \mu_S(S) \frac{1}{(2\delta)2^n} \sum_{x\in S} 1_{f(x)=C(x)} \tag{2}$$

$$= \frac{1}{(2\delta)2^n} \sum_{x\in\{0,1\}^n} 1_{f(x)=C(x)} \sum_{S\ni x} \mu_S(S) \tag{3}$$

$$= \frac{1}{(2\delta)2^n} \sum_{x\in\{0,1\}^n} 1_{f(x)=C(x)} Pr_{S\sim\mu_s}[x \in S] \tag{4}$$

$$= \frac{1}{(2\delta)2^n} \sum_{x\in\{0,1\}^n} \nu(x) 1_{f(x)=C(x)}. \tag{5}$$

Equations 1 and 5, combined with the fact that we are in Case 1, imply $\mathbb{E}[Y] = \mathbb{E}_{S\sim\mu_s}[M_{S,C}](2\delta)2^n \leq (1/2+\epsilon)(2\delta)2^n$. Note that Y is a sum of independent indicator variables, so by the multiplicative Chernoff bound and $\epsilon < 1/2$

$$\begin{aligned}\Pr[Y > (1/2+2\epsilon)(2\delta)2^n] &= \Pr[Y > (1/2+\epsilon)(2\delta)2^n + \epsilon(2\delta)2^n] \\ &\leq e^{-\frac{1}{3}\left(\frac{\epsilon}{.5+\epsilon}\right)^2 (2\delta)2^n} \leq e^{-\frac{1}{3}\epsilon^2(2\delta)2^n}.\end{aligned}$$

This was for a fixed C. By the union bound, with probability at least

$$1 - (\text{number of Small circuits}) e^{-\frac{1}{3}\epsilon^2(2\delta)2^n} - e^{-\eta^2(2\delta)2^n/3} \tag{6}$$

$$= 1 - (s')^{o(s')} e^{-\frac{1}{3}\epsilon^2(2\delta)2^n} - e^{-\eta^2(2\delta)2^n/3} \tag{7}$$

we have that for all Small circuits, the number of agreements between C and f on H is at most $(1/2+2\epsilon)(2\delta)2^n$ and $|H| \geq (1-\eta)(2\delta)2^n$. Provided 7 is positive, there exists such an H, so we can choose $s' = \epsilon^2\delta s/5n \leq 2^{.9n}\epsilon^2\delta$ and $\eta = 5(\delta 2^n)^{-1/2}$ provided n is large enough.

0.2 Case 2:

Now suppose we are in Case 2. We want a small circuit that agrees with f on a $(1-\delta)$ fraction of inputs. First sample $C_1 \ldots C_t$ from μ_C.. Let $\eta(x) = \Pr_{C\in\mu_C}[C(x) = f(x)]$. Recall that $M_{S,C} = Pr_{x\in S}[C(x) = f(x)] = \mathbb{E}_{x\in S}[1_{C(x)=f(x)}]$. So for all Large sets S,

$$\begin{aligned}\mathbb{E}_{C\in\mu_C}[M_{S,C}] = \mathbb{E}_{C\in\mu_C}\mathbb{E}_{x\in S}[1_{C(x)=f(x)}] = \mathbb{E}_{x\in S}\mathbb{E}_{C\in\mu_C}[1_{C(x)=f(x)}] = \\ \mathbb{E}_{x\in S}\eta(x) \geq 1/2+\epsilon.\end{aligned} \tag{8}$$

Let $n'(x) = \frac{1}{t}\sum_{i=1}^t 1_{C_i(x)=f(x)}$.

Claim 5. *With probability at least* $1 - 2^n e^{-.5\epsilon^2 t}$ *over choice of* $C_1, \ldots C_t$*, for all* x*,* $|\eta(x) - \eta'(x)| < \epsilon/2$.

Proof. Use Chernoff to bound the probability of the condition not holding for a specific x: by the additive form of Chernoff, $Pr[|\eta'(x) - \eta(x)| \geq \epsilon/2] \leq e^{-.5\epsilon^2 t}$. Next union bound over $x \in \{0,1\}^n$. $\square$

If $t \geq 3n/\epsilon^2$, there are $C_1, \ldots C_t$ such that for all x, $|\eta(x) - \eta'(x)| < \epsilon/2$. By 8, $\mathbb{E}_{x \in S}[\eta'(x)] \geq 1/2 + \epsilon/2$ for all S of size $(2\delta)2^n$.

A first attempt: Define $C^*(x) = Maj(C_1(x), \ldots C_t(x))$. Let S^* be the Large set where $\mathbb{E}_{x \in S}[\eta'(x)]$ is the smallest. In other words, S^* is the $(2\delta)2^n$ first x in increasing order of $\eta'(x)$. Note that for $x \notin S^*$, $n'(x) \geq 1/2 + \epsilon/2$. This implies C^* is correct outside S^*, since C^* is the majority of the C_i and at least half the C_i are correct. So $Pr_{x \in \{0,1\}^n}[C^*(x) = f(x)] \geq 1 - 2\delta$, but we need $1 - \delta$. It seems like this approach won't tell us much about what happens inside S^*. To this end, we modify C^*.

The real attempt: Let S^* be the defined as in the first attempt. Define $\beta = \max\{\eta'(x) : x \in S^*\} - 1/2$. Define

$$C^*(x) = \begin{cases} 0 & \frac{1}{t}\sum C_i(x) \leq \frac{1}{2} - \beta \\ 1 & \frac{1}{t}\sum C_i(x) > \frac{1}{2} + \beta \\ Y & \text{otherwise,} \end{cases} \tag{9}$$

where Y is chosen independently in $\{0,1\}$ with mean $\frac{1}{2} + \frac{\frac{1}{t}\sum_i C_i(x) - \frac{1}{2}}{2\beta}$. For $x \notin S^*$, $\eta'(x) \geq 1/2 + \beta > 1/2$ by the definition of β, so C^* is correct on x. Within S^*, we need to show that the probability over choice of x and Y that $C^*(x)$ is correct is at least $1/2$. If $f(x) = 1$, then $\frac{1}{t}\sum C_i(x) = \eta'(x)$ and so

$$\begin{aligned} \Pr_{x \in S^*, Y}[C^*(x) = f(x)] &= \mathbb{E}_{x \in S^*}[\mathbb{E}_Y[C^*(x)]] \\ &\geq E_{x \in S^*}\left[\frac{1}{2} + \frac{\frac{1}{t}\sum_i C_i(x) - \frac{1}{2}}{2\beta}\right] \\ = E_{x \in S^*}\left[\frac{1}{2} + \frac{\eta'(x) - \frac{1}{2}}{2\beta}\right] &\geq 1/2 + \epsilon/4\beta > 1/2. \end{aligned}$$

because even if $\frac{1}{t}\sum_i C_i(x) \leq 1/2 - \beta$, $C^*(x)$ will be zero, which is larger than the negative number $\frac{1}{2} + \frac{\frac{1}{t}\sum_i C_i(x) - \frac{1}{2}}{\beta}$. If $f(x) = 0$, then $\frac{1}{t}\sum C_i(x) = 1 - \eta'(x)$ and so

$$\begin{aligned} \Pr_{x,Y}[C^*(x) = f(x)] &= \mathbb{E}_x[\mathbb{E}_Y[1 - C^*(x)]] \\ &\geq E_{x \in S^*}\left[\frac{1}{2} + \frac{1/2 - \frac{1}{t}\sum_i C_i(x)}{2\beta}\right] \\ = E_{x \in S^*}\left[\frac{1}{2} + \frac{\eta'(x) - \frac{1}{2}}{2\beta}\right] &\geq 1/2 + \epsilon/4\beta > 1/2. \end{aligned}$$

If remains to show that C^* can be computed by a small circuit. This stands to reason, since in addition to the t copies of C we just need to add a gadget that outputs 1 if the sum of the C_i is large enough, 0 if it is small enough, and a random bit with the right probability if it is in between. Intuitively this should take something like $O(t + ts')$ gates. First, let's show C^* can be computed by a small randomized circuit $C(x, r)$ depending on some random bits r. As 2β is q/t for some integer $q \in [t]$, look at

$$\frac{1}{2} + \frac{\frac{1}{t}\sum_i C_i(x) - \frac{1}{2}}{2\beta} \tag{10}$$

$$= \frac{1}{2q}(q + 2\sum_i C_i(x) - t). \tag{11}$$

This means once we can compute $q + 2\sum_i C_i(x) - t$ from the C_i, we can output 0 if it is ≤ 0 and output 1 if it is $> 2q$, and if it is in $[2q]$ output 1 or 0 according to whether it is $\geq$ or $<$ a random binary number between 1 and $2q$ generated by at most $O(\log t)$ random bits. Since $q+2\sum_i C_i(x)-t$ is at most $2t$, we need at most $O(\log t)$ bits to represent it and $O(t)$ gates to compute each bit (including a sign bit). To compare the two numbers we only need $O(\log^2 t)$ gates. In any case, since $t = 3n/\epsilon^2$, we need at most

$$3ns'/\epsilon^2 + O(n/\epsilon^2 \log(n/\epsilon^2) \leq s$$

gates.

If $C(x, r)$ is a randomized circuit (where x is the input and r is random bits) such that $E_x[E_r[1_{f(x)=C(x,r)}]] \geq 1 - \delta$, then there exists r_0 such that $E_x[1_{f(x)=C(x,r_0)}] \geq 1 - \delta$. Fix that r_0 and hardwire it into C. This gives a circuit of size at most s that computes f with probability at least $1 - \delta$.

□