CSE 533: The PCP Theorem and Hardness of Approximation **(Autumn 2005)**

Lecture 12: "Confuse/Match" Games (I)

Nov. 9, 2005

Lecturer: Ryan O'Donnell *Scribe: Ning Chen*

1 Confuse/Match Games

Feige and Kilian [1] showed something similar to the Parallel Repetition Theorem if the 2P1R game G has some extra property, as the following theorem shows.

Theorem 1.1. *Let $s < 1$ be a constant. Suppose G is a "confuse/match"-style game with $\omega(G) \leq s$. Then if $k = \mathrm{poly}(1/\epsilon)$, $\omega(G^k) < \epsilon$.*

We will prove this theorem by the end of the next lecture.

Definition 1.2. (Confuse/Match Game) *Let G be any 2P1R game with the projection property. Then the* confuse/match *version, denoted by $G_{c/m}$, works as follows:*

With probability 50%, the verifier plays the original game G (this is a "match round").
With probability 50%, the verifier:

- *picks two sets of questions independently, (x_1, y_1) and (x_2, y_2)*
- *sends x_1 to P_1, sends y_2 to P_2*
- *always accepts*

(this is a "confuse round").

Remarks:

- The "confuse" part is indeed confusing to the provers, in that P_1 and P_2 see their normal probability distribution on questions, and cannot tell if the verifier is doing "match" or "confuse".
- It is easy to see that if $\omega(G) = 1 - \delta$, then $\omega(G_{c/m}) = 1 - \delta/2$.
- If the game is played only once, using the confuse/match version of the game is pointless. It only helps if the game is played with parallel repetitions.
- Despite the fact that we consider only games G with the "projection property", the confuse/match version $G_{c/m}$ will *not* have the projection property (so neither will the parallel repeated version). Feige and Kilian [1] fixed this via a new style game called miss/match game: Instead of "confuse" part, we have a "miss" part as follows: The verifier

– picks (x, y) as usual,
– sends x to P_1,
– sends "miss" to P_2,
– always accepts assuming P_2's answer is "miss".

Miss/match games indeed have the projection property, and Theorem 1.1 holds for them as well (with 99.5% of the proof details being the same). For simplicity, we will just focus on confuse/match games though.

Intuition for Theorem 1.1. The intuition for the theorem is roughly as follows. Focus on P_2's strategy. On the one hand (roughly speaking), this strategy could be "mostly serial", meaning that the answers it gives in the ith coordinate more or less only depends on the question it gets in the ith coordinate. In this case, P_2 is not taking advantage of the fact that it gets to see all its questions simultaneously, and it will win on all coordinates only with exponentially small probability. On the other hand (roughly speaking), P_2's answers could strongly depend on many question-coordinates simultaneously. However, for any confuse-round question P_2 bases an answer on, P_1 has no information about what P_2 sees. Thus it is very hard for P_1 to coordinate with P_2 given such a strategy.

Indeed, most of the proof of Theorem 1.1 is devoted to making a dichotomy statement like this about strategies rigorous. Having done that, the proof that the provers win with low probability is quite easy.

In fact, this dichotomy theorem and the same intuition equally well explains why the Theorem holds for miss/match games. In this case, focus on P_1's strategy. If it is mostly serial, the players will win only with exponentially small probability. Otherwise, answers are based on the questions in many coordinates — and on many of these coordinates P_2 only sees "miss". In other words, it's really only necessary that *one* of the provers be "confused". (In confuse/match games, *both* get confused.)

2 Revealing a single random question changes little — a lemma

Lemma 2.1. *Let X be a set and γ a probability distribution on X. Let $f : X^C \to \{0, 1\}$, $c \geq 1$, where we think of X^C as having the product probability distribution γ^C. Let*

$$\mu = \mathop{\mathbf{E}}_{\vec{x} \in \gamma^c}[f(\vec{x})] = \mathop{\mathbf{Pr}}_{\vec{x} \in \gamma^c}[f(\vec{x}) = 1].$$

Suppose we pick $i \in \{1, \ldots, c\}$ uniformly at randomly and pick $x_i \leftarrow \gamma$ at random. Let

$$\widetilde{\mu} = \widetilde{\mu}_{i,x_i} = \mathop{\mathbf{E}}_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_c}[f \mid x_i].$$

Then we have

(a)

$$0 \leq \mathop{\mathbf{E}}_{i,x_i}\left[(\widetilde{\mu}-\mu)^2\right] \leq \mu/C \leq 1/C$$

(a')

$$\mathop{\mathbf{E}}_{i,x_i}\left[(\widetilde{\mu}-\mu)^2\right] = \mathop{\mathbf{E}}_{i,x_i}\left[(\widetilde{\mu})^2\right] - \mu^2$$

(b)

$$\mathop{\mathbf{Pr}}_{i,x_i}\left[|\widetilde{\mu}-\mu| \geq \delta\right] \leq \delta, \quad \textit{where } \delta = 1/\sqrt[3]{C}$$

Plausibility argument. At one extreme, f's value might depend on many coordinates (imagine $X = \{0,1\}$ and f is the Majority function); then knowing one coordinate does not make much difference. At the other extreme, f's value might only depend on a single coordinate (imagine $f(\vec{x}) = x_i$); but then the probability that we pick i to be this coordinate is only $1/C$.

The rest of this section is devoted to the proof of this lemma.

For part (a'), we have

$$\begin{aligned}
\mathop{\mathbf{E}}_{i,x_i}\left[(\widetilde{\mu}-\mu)^2\right] &= \mathop{\mathbf{E}}_{i,x_i}\left[(\widetilde{\mu})^2\right] - 2\mu \mathop{\mathbf{E}}_{i,x_i}[\widetilde{\mu}] + \mu^2 \\
&= \mathbf{E}\left[(\widetilde{\mu})^2\right] - 2\mu \mathop{\mathbf{E}}_{\vec{x}}[f(\vec{x})] + \mu^2 \\
&= \mathbf{E}\left[(\widetilde{\mu})^2\right] - 2\mu^2 + \mu^2 \\
&= \mathbf{E}\left[(\widetilde{\mu})^2\right] - \mu^2.
\end{aligned}$$

Part (b) follow easily from part (a):

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{i,x_i}\left[|\widetilde{\mu}-\mu| \geq 1/\sqrt[3]{C}\right] &= \mathbf{Pr}\left[(\widetilde{\mu}-\mu)^2 \geq 1/C^{2/3}\right] \\
&\leq \frac{\mathbf{E}[(\widetilde{\mu}-\mu)^2]}{1/C^{2/3}} \\
&\leq \frac{1/C}{1/C^{2/3}} \\
&= 1/\sqrt[3]{C},
\end{aligned}$$

where the first inequality is by Markov inequality and the second one is by part (a).

So it remains to prove part (a). For any $i \in \{1,\ldots,C\}$, define $f^i : X \to \mathbb{R}$, where

$$f^i(x_i) = \mathop{\mathbf{E}}_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_c}[f \mid x_i] - \mu.$$

By abuse of notation, we will also write $f^i : X^C \to \mathbb{R}$, where $f^i(\vec{x}) = f^i(x_i)$. Let us also define $W_i = \mathbf{E}_{x_i}\left[f^i(x_i)^2\right]$. Thus, we have

$$\begin{aligned}
\mathop{\mathbf{E}}_{i,x_i}\left[(\widetilde{\mu}-\mu)^2\right] &= \mathop{\mathbf{E}}_{i}\left[\mathop{\mathbf{E}}_{x_i}[(\widetilde{\mu}-\mu)^2]\right] \\
&= \mathop{\mathbf{E}}_{i}\left[\mathop{\mathbf{E}}_{x_i}[f^i(x_i)^2]\right] \\
&= \frac{1}{C}\sum_{i=1}^{C}\mathop{\mathbf{E}}_{x_i}\left[f^i(x_i)^2\right] \\
&= \frac{1}{C}\sum_{i} W_i.
\end{aligned}$$

Hence all we need to show is $\sum_i^C W_i \le \mu$.

Fact 2.2.

$$\mathop{\mathbf{E}}_{x_i}\left[f^i(x_i)\right] = 0.$$

Proof.

$$\mathop{\mathbf{E}}_{x_i}\left[f^i(x_i)\right] = \mathop{\mathbf{E}}_{x_i}\left[\mathop{\mathbf{E}}_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_c}[f \mid x_i] - \mu\right] = \mathbf{E}[f] - \mu = 0.$$

□

Fact 2.3. *If $i \neq j$, then*

$$\mathop{\mathbf{E}}_{\vec{x}}\left[f^i(\vec{x})f^j(\vec{x})\right] = 0.$$

Proof.

$$\begin{aligned}
\mathop{\mathbf{E}}_{\vec{x}}\left[f^i(\vec{x})f^j(\vec{x})\right] &= \mathop{\mathbf{E}}_{x_i,x_j}\left[f^i(x_i)f^j(x_j)\right] && (f^\ell \text{ depends only on } x_\ell) \\
&= \mathop{\mathbf{E}}_{x_i}\left[f^i(x_i)\right]\cdot\mathop{\mathbf{E}}_{x_j}\left[f^j(x_j)\right] && (x_i \text{ and } x_j \text{ are independent}) \\
&= 0 && \text{(Fact 2.2)},
\end{aligned}$$

□

Finally, let us define $g : X^c \to \mathbb{R}$ by $g(\vec{x}) = \sum_{i=1}^C f^i(\vec{x})$. Thus,

$$0 \le \mathop{\mathbf{E}}_{\vec{x}}\left[(f(\vec{x}) - g(\vec{x}))^2\right] = \mathop{\mathbf{E}}_{\vec{x}}\left[(f(\vec{x}))^2\right] - 2\mathop{\mathbf{E}}_{\vec{x}}\left[f(\vec{x})g(\vec{x})\right] + \mathop{\mathbf{E}}_{\vec{x}}\left[(g(\vec{x}))^2\right]$$

Note that

1.
$$\mathop{\mathbf{E}}_{\vec{x}}\left[(f(\vec{x}))^2\right] = \mu \qquad \text{(since } f \text{ is 0-1 valued)}$$

2.

$$\begin{aligned}
E_{\vec{x}}\left[f(\vec{x})g(\vec{x})\right] &= \sum_i \mathop{\mathbf{E}}_{\vec{x}}\left[f(\vec{x})f^i(\vec{x})\right] \\
&= \sum_i \mathop{\mathbf{E}}_{x_i}\mathop{\mathbf{E}}_{x_j:j\neq i}\left[f(x_i;x_j\text{'s})f^i(x_i)\right] \\
&= \sum_i \mathop{\mathbf{E}}_{x_i}\left[f^i(x_i)\mathop{\mathbf{E}}_{x_j:j\neq i}[f(x_i;x_j\text{'s})]\right] \\
&= \sum_i \mathop{\mathbf{E}}_{x_i}\left[f^i(x_i)(f^i(x_i)+\mu)\right] \qquad \text{(by definition of } f^i\text{)} \\
&= \sum_i W_i + \sum_i \mu\cdot\mathbf{E}[f^i] \\
&= \sum_i W_i \qquad \text{(by Fact 2.2)}
\end{aligned}$$

3.

$$\mathop{\mathbf{E}}_{\vec{x}}\left[(g(\vec{x}))^2\right] = \sum_{i,j}\mathbf{E}\left[f^i(\vec{x})f^j(\vec{x})\right] = \sum_{i=1} W_i \quad \text{(by Fact 2.3)}$$

Putting the above three equalities together, we have $0 \le \mu - 2\sum_i W_i + \sum_i W_i = \mu - \sum_i W_i$. This completes the proof of the lemma.

3 Predictability

Definition 3.1. *Let $P : Q^c \to A^c$ be a prover strategy, where Q is the set of questions and A is the set of answers; let γ^C be a product distribution on Q^C as before. Let $R \subseteq \{1,\ldots,C\}$, $R \neq \emptyset$. Define the predictability as follows*

$$\text{Predictability}_R(P) \triangleq \sum_{\vec{a}\in A^R}\left(\mathop{\mathbf{Pr}}_{\vec{q}\in\gamma^C}[P(\vec{q})[R]=\vec{a}]^2\right).$$

Here are some observations

- If there were no square in the definition, the sum would always be 1.
- $0 <$ Predictability$_R(P) \le 1$.
- If $P(\vec{q})[R]$ is completely determined, then Predictability$_R(P) = 1$.
- If $P(\vec{q})[R]$ is uniformly distributed on some set of size N, then Predictability$_R(P) = 1/N$.
- If $i \notin R$, then Predictability$_{R\cup\{i\}}(P) \le$ Predictability$_R(P)$, and the equality holds if and only if the answers on R force the answer on i.

From Theorem 1.1, we have the following rather easily:

Corollary 3.2. *Let $R \neq \emptyset$ be any set, and $P : Q^{C'} \to A^C$. Suppose i and q_i are chosen randomly as in Theorem 1.1; then,*

$$\mathop{\mathbf{E}}_{i,q_i}[\text{Predictability}_R(P \mid i, q_i)] - \text{Predictability}_R(P) \leq 1/C'.$$

Proof. For each $\vec{a} \in A^R$, let $f_{\vec{a}} : Q^{C'} \to \{0, 1\}$ be the indicator that $P(\vec{q})[R] = \vec{a}$. Let

$$\mu_{\vec{a}} = \mathop{\mathbf{E}}_{\vec{q}}[f_{\vec{a}}(\vec{q})] = \mathop{\mathbf{Pr}}_{\vec{q}}[P(\vec{q})[R] = \vec{a}],$$

and

$$\widetilde{\mu}_{\vec{a},i,q_i} = \mathbf{E}[f_{\vec{a}} \mid i, q_i].$$

Due to Theorem 1.1 (parts (a) and (a')), for any $\vec{a}$, we have

$$\mathop{\mathbf{E}}_{i,q_i}\left[(\widetilde{\mu}_{\vec{a},i,q_i})^2\right] - \mu_{\vec{a}}^2 \leq \frac{\mu_{\vec{a}}}{C'}.$$

Sum over all $\vec{a}$, and notice that $\sum_{\vec{a}} \mu_{\vec{a}} = 1$; we are done. □

References

[1] U. Feige, J. Kilian, Two prover protocols: low error at affordable rates, STOC 1994, 172-183.

[2] R. Raz, A parallel repetition theorem, STOC 1995, 447-456.