CSE533: Information Theory in Computer Science November 1, 2010

Lecture 10 + additional notes

Lecturer: Anup Rao *Scribe: Mohammad Moharrami*

1 Constraint satisfaction problems

We start by defining bivariate constraint satisfaction problems. Such a problem $\mathcal{C}$ is specified by a bipartite graph with n vertices on each side, edge set $E \subseteq [n] \times [n]$, and a family of constraints $\{C_{i,j}\}_{(i,j)\in E}$, with $C_{i,j} \subseteq [k] \times [k]$. k is called the *alphabet size.* Throughout this lecture, we treat k as a constant. Given such a problem, we define its value:

$$\mathrm{val}(\mathcal{C}) \stackrel{\mathrm{def}}{=} \max_{a,b\in[k]^n} \Pr_{(i,j)\in E}[(a_i, b_j) \in C_{ij}],$$

where the probability is over a uniformly random edge of the graph, and the maximum is over assignments a, b to the vertices of the graph. In words, the value is the maximum fraction of constraints that can be satisfied by any assignment to the vertices of the graph.

1.1 Example: 3-SAT

Let Φ be a 3-SAT formula. Let n be larger than both the number of variables and the number of clauses. We construct a bivariate constraint satisfaction problem from Φ as follows. Let (i, j) be an edge in $E \subseteq [n] \times [n]$ if and only if the i'th clause contains the j'th variable. Set $k = 7$. Think of every assignment $a_i \in [7]$ to a vertex on the left as representing one of the 7 satisfying assignments to the i'th clause, and every assignment $b_j \in [2]$ as corresponding to one of two boolean assignments to the j'th variable. For each edge (i, j), let $C_{i,j}$ be the set of assignments to the vertices i, j where the j'th variable gets the value b_j in the assignment a_i.

If Φ, is satisfiable, the value of the corresponding problem is 1. On the other hand, if the value of the problem is 1, then the maximizing assignment a, b gives a satisfying assignment b for the variables in Φ. Thus, determining whether $\mathrm{val}(\mathcal{C}) = 1$ or not is NP-hard, and in fact, NP-complete.

1.2 Probabilistically Checkable Proofs (PCPs) and Parallel Repetition

The *PCP Theorem* of Arora et al [1] shows that if $P \neq NP$, there are no polynomial time approximation algorithms that can guarantee an arbitrarily close approximation to $\mathrm{val}(\mathcal{C})$.

Theorem 1 (PCP Theorem [1]). *There exists a positive constant α, an alphabet size k, and a polynomial time computable function f mapping boolean formulas to constraint satisfaction problem instances with alphabet size k, such that*

$$\mathrm{val}(f(\Phi)) \text{ is } \begin{cases} 1 & \text{if } \Phi \text{ is satisfiable,} \\ < 1 - \alpha & \text{if not.} \end{cases}$$

If one had a polynomial time algorithm that could distinguish $\mathcal{C}$ with $\mathrm{val}(\mathcal{C}) = 1$ from value $\mathrm{val}(\mathcal{C}) < 1 - \alpha$, one could use f from the PCP theorem to solve SAT. Thus, the theorem proves an amazing qualitative statement: even approximating the value of bivariate constraint satisfaction problems is NP-hard. Raz's parallel repetition theorem implies the following quantitative strengthening of the theorem:

Theorem 2 (PCP Theorem + Parallel Repetition)**.** *For every $\varepsilon > 0$, there is an alphabet size k, and a polynomial time computable function f mapping boolean formulas to constraint satisfaction problem instances with alphabet size k, such that*

$$\mathrm{val}(f(\Phi)) \textit{ is } \begin{cases} 1 & \textit{if } \Phi \textit{ is satisfiable,} \\ < \varepsilon & \textit{if not.} \end{cases}$$

A natural way to try and prove Theorem 2 from Theorem 1 is to find a polynomial time computable transformation to turn $\mathcal{C}$ into $\mathcal{C}'$ such that if $\mathrm{val}(\mathcal{C}) = 1$, then $\mathrm{val}(\mathcal{C}') = 1$, but if $\mathrm{val}(\mathcal{C}) < 1 - \alpha$, then $\mathrm{val}(\mathcal{C}') < \varepsilon$. Perhaps the first such transformation one might think of is to repeat the problem in parallel. Given $\mathcal{C}$ as above, we define $\mathcal{C}^t$ as follows. The vertex set on each side is the t-fold cartesian product $[n]^t$. $\bar{i} = (i_1, \ldots, i_t)$ is connected to $\bar{j} = (j_1, \ldots, j_t)$ in the edge set of $\mathcal{C}^t$ if and only if (i_r, j_r) is an edge of $\mathcal{C}$ for each r. Thus, each new edge corresponds to an element of E^t. Finally, we define the constraint $C_{\bar{i},\bar{j}} = \{(x, y) \in [k]^t \times [k]^t : \text{ for all } r,\ (x_r, y_r) \in C_{i_r, j_r}\}$.

1.3 2-Player Games

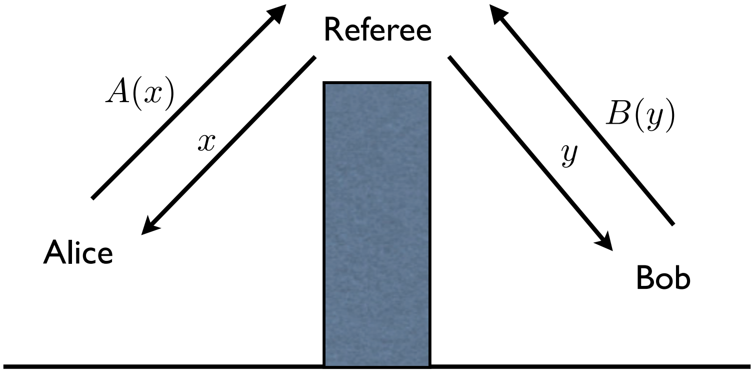

Figure 1: 2 Player Game.

Another way to view bivariate constraint satisfaction problems that will be convenient for us, is in terms of a particular kind of 2 player game. The game is played by two cooperating players, Alice and Bob and is officiated by a referee. Alice and Bob are not allowed to talk during the run of the game. The referee samples a uniformly random edge X, Y from the edge set of $\mathcal{C}$ and sends X to Alice and Y to Bob. Each player then responds with an assignment to the vertex that they see $(A(X), B(Y))$. They win the game if and only if $(A(X), B(Y)) \in C_{X,Y}$.

Observe that any strategy for playing the above game gives an assignment to the vertices, and vice versa. Thus, $\mathrm{val}(\mathcal{C})$ is the probability that Alice and Bob win the game using the best strategy. Notice, that we can also assume that Alice and Bob have access to a shared random string that they can use to generate their answers, since any strategy that uses such a string is simply a convex

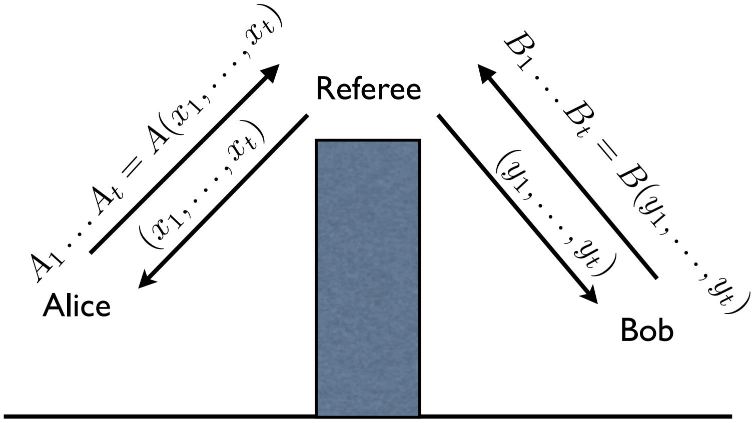
$A_1 \dots A_t = A(x_1, \dots, x_t)$
$(x_1, \dots, x_t)$
Referee
Alice
$(y_1, \dots, y_t)$
$B_1 \dots B_t = B(y_1, \dots, y_t)$
Bob

Figure 2: The repeated game

combination of strategies that are deterministic. So allowing these strategies does not change the value of the game.

In the parallel version of the game, the referee samples t uniformly random edges $(X_1, Y_1) \dots (X_t, Y_t)$ and sends $(X_1 \dots X_t)$ to Alice and $(Y_1 \dots Y_t)$ to Bob. They respond with answers $A_1 \dots A_t = A(X_1, \dots, X_t))$ from Alice and $B_1 \dots B_t = B(Y_1, \dots, Y_t)$ from Bob. They win if and only if they win in each coordinate, namely if for all r, $(A_r, B_r) \in C_{X_r, Y_r}$.

2 The value of repeated games

Clearly, $\text{val}(\mathcal{C})^t \leq \text{val}(\mathcal{C}^t)$, since Alice and Bob could use their best strategy for $\mathcal{C}$ in each of the coordinates. One might hope that $\text{val}(\mathcal{C})^t \leq \text{val}(\mathcal{C}^t)$. If that was true, we would arrive at Theorem 2 from Theorem 1, since $\mathcal{C}^t$ is easily computable from $\mathcal{C}$ in polynomial time. Indeed, this strategy will eventually succeed.

However, as the following example shows, strict equality does not always hold:

2.1 A counterexample

Let X, Y be uniformly random bits. The players win the game if and only if either $A = B = (1, X)$ or $A = B = (2, Y)$. Thus, the players win only if they both decide to guess the same input bit and succeed in doing so. It is easy to check that the value of this game is $1/2$ — if they win when $(X = 0, Y = 0)$ and $(X = 0, Y = 1)$, Bob must be trying to guess Alice's bit on every input, so he must guess incorrectly whenever $X = 1$. This means they can win on at most two pairs of questions, no matter what. We show that $\text{val}(\mathcal{C}^2) = 1/2$, using the following strategy:

$$\begin{array}{rr} A_1 = (1, X_1) & A_2 = (2, X_1) \\ B_1 = (1, Y_2) & B_2 = (2, Y_2). \end{array}$$

Clearly, they win the parallel repetition whenever $X_1 = Y_2$, which happens with probability $1/2$. To understand what is going on in this example a little better, let us define the events W_1, W_2 to be the event of winning the game in each coordinate, with the above answer strategy. Then observe that:

$$\Pr[W_1 \wedge W_2] = \Pr[W_1] \cdot \Pr[W_2|W_1] = (1/2) \cdot 1.$$

The first term in this product behaves as we expect. The second term is more interesting. Conditioned on W_1, we have that $Y_2 = X_1$ with probability 1! Thus, in this conditioned space, Alice knows the value of Y_2: it is equal to X_1, which she knows. In some sense, this strategy for the players allows them to communicate, at least in the probability space obtained by conditioning on their winning in the first coordinate. As we shall see, this communication is the major obstacle that needs to be overcome in order to prove strong versions of the parallel repetition theorem.

2.2 Parallel repetition theorems

2.2.1 A weak bound

The first upper bound on $\mathrm{val}(\mathcal{C}^t)$ was obtained by Verbitsky, based on the density version of the Hales-Jewett theorem. Here we sketch the connection. The Hales-Jewett theorem concerns structures called combinatorial lines. A set $L \subseteq [m]^t$ of size m is called a combinatorial line if there exists a subset $K \subseteq [t]$, such that for every $x, y \in L$ and $i, j \in K, r \notin K$, we have $x_i = x_j$ and $x_r = y_r$. It is a set of m points where the coordinates in K are equal and range over all m values, and the other coordinates are fixed.

Theorem 3 ([2])**.** *For every m, ε there exists a constant t_0 such that if $t \geq t_0$ and $S \subseteq [m]^t$ is of size at least $\varepsilon \cdot m^t$, then S contains a combinatorial line.*

Verbitsky observed that this theorem implies the following theorem:

Theorem 4 ([5])**.** *If $\mathrm{val}(\mathcal{C}) < 1$, then $\lim_{t\to\infty} \mathrm{val}(\mathcal{C}^t) = 0$.*

Theorem 4 easily follows from Theorem 3. Suppose for the sake of contradiction that $\mathrm{val}(\mathcal{C}^t)$ is at least ε for an infinite number of settings of t. Let E be the edge set of $\mathcal{C}$, $m = |E|$ and let t_0 be as in Theorem 3. Then given any $t > t_0$ and any strategy for playing $\mathcal{C}^t$, let $S \subseteq E^t$ be the set of edges of $\mathcal{C}^t$ whose constraints are satisfied by the strategy. We claim that $|S| < \varepsilon$. If not, then by Theorem 3, S must contain a combinatorial line L. Alice and Bob can use L to win $\mathcal{C}$ with probability 1 — on question x, y, they can each embed their questions into the unfixed coordinates of L to obtain t pairs of questions for $\mathcal{C}^t$ that is an element of L. This can be done without communicating. They answer according to the strategy for $\mathcal{C}^t$ to win with probability 1, contradicting the fact that $\mathrm{val}(\mathcal{C}) < 1$.

Theorem 4 is *not* strong enough to prove Theorem 2, since t_0 (necessarily) depends on n. Theorem 4 leaves open the possibility that (say) $\exp(n)$ repetitions are necessary (indeed the dependence of t_0 on m is a tower of exponentials) to reduce the value of $\mathcal{C}$ to less than ε, which means that the time to compute $\mathcal{C}^{t_0}$ would be superpolynomial in n.

2.3 A strong bound

Raz [4] proved the following theorem:

Theorem 5. *If $\mathrm{val}(\mathcal{C}) < 1 - \alpha$, and k is the alphabet size, then $\mathrm{val}(\mathcal{C}^t) = 2^{-\Omega_{\alpha,k}(t)}$.*

Here we present a simplified version of his proof due to Holenstein [3].

Fix a strategy for the repeated game $\mathcal{C}^t$. Let

$$\begin{matrix} X^n \\ Y^n \end{matrix} = \begin{matrix} X_1 X_2 \cdots X_t \\ Y_1 Y_2 \cdots Y_t \end{matrix},$$

be the random variables of the questions in the game, and let

$$\begin{matrix} A^n \\ B^n \end{matrix} = \begin{matrix} A_1 A_2 \cdots A_t \\ B_1 B_2 \cdots B_t \end{matrix},$$

be the corresponding answers.

Given any set $S \subseteq [t]$, let W_S denote the event of winning in the games corresponding to the coordinates of S. The proof follows from this lemma:

Lemma 6. *There exists a constant $\gamma(k, \alpha)$ such that for all $S \subseteq [t]$, with $|S| < \gamma t$, either $\Pr[W_S] < 2^{-\gamma t}$ or there exists an index i such that $\Pr[W_{\{i\}} | W_S] < 1 - \gamma$.*

Let us first show how to use Lemma 6 to prove Theorem 5. Suppose the value of the game is at least $2^{-\gamma t}$. Then, by repeatedly applying Lemma 6, we get a sequence of indices $i_1, \ldots, i_{\gamma t}$, such that for each ℓ, $\Pr[W_{\{i_\ell\}} | W_{\{i_1, \ldots, i_{\ell-1}\}}] < 1 - \gamma$. Thus, $\Pr[W_{\{i_1, \ldots, i_\ell\}}] < (1-\gamma)^{\gamma t}$ and so we have argued that for any fixed strategy, $\Pr[W_{[n]}] \leq \max\{2^{-\gamma t}, (1-\gamma)^{\gamma t}\}$, proving Theorem 5.

In order to prove Lemma 6, we shall need the following facts about divergence, alluded to in earlier lectures. The first says that an event with probability 2^{-d} can change a distribution by at most d in terms of its divergence:

Lemma 7. *Let E be an event. Then*

$$D(p(x|E) \| p(x)) \leq \log(1/p(E)).$$

Convexity arguments can be used to prove the following mild generalizations of Lemma 7:

Lemma 8. *Let E be an event and A, X be random variables such that the support of A is of size k. Then,*

$$\mathop{\mathbb{E}}_{A|E} [D(p(x|E, A) \| p(x))] \leq \log(k/p(E)).$$

Proof

$$\begin{aligned} & \mathop{\mathbb{E}}_{A|E} [D(p(x|A, E) \| p(x))] \\ & \leq \mathop{\mathbb{E}}_{A|E} [\log(1/p(A, E))] && \text{by Lemma 7} \\ & \leq \log\left(\mathop{\mathbb{E}}_{A|E} [1/p(A, E)]\right) && \text{by concavity of log} \\ & = \log\left(\sum_a 1/p(E)\right) \\ & = \log(k/p(E)). \end{aligned}$$

■

Lemma 9. *Let E be an event and U, A, X be random variables such that the support of A is of size k. Then*

$$\mathop{\mathbb{E}}_{A,U|E}\left[D(p(x|E,A,U)\|p(x|U))\right] \leq \log(k/p(E)).$$

Proof

$$\begin{aligned}
&\mathop{\mathbb{E}}_{A,U|E}\left[D(p(x|E,A,U)\|p(x|U))\right] \\
&\leq \mathop{\mathbb{E}}_{U|E}\left[\log(k/p(E|U))\right] && \text{by Lemma 7} \\
&\leq \log\left(k \cdot \mathop{\mathbb{E}}_{U|E}\left[1/p(E|U)\right]\right) && \text{by concavity of log} \\
&\leq \log\left(k \cdot \sum_u p(u)/p(E)\right) \\
&= \log(k/p(E)).
\end{aligned}$$

■

The next lemma says that divergence adds if the base distribution is a product distribution.

Lemma 10. *Let $p(x, y)$ and $q(x, y) = q(x)q(y)$ be two distributions. Then*

$$D\big(p(x,y)\|q(x,y))\big) \geq D\big(p(x)\|q(x)\big) + D\big(p(y)\|q(y)\big).$$

The final lemma says that if the divergence between two distributions is small, then so is the statistical distance:

Lemma 11. $D(p\|q) \geq |p - q|^2$.

Proof of Lemma 6 For a parameter γ that we shall set later, let us assume that $\Pr[W_S] \geq 2^{-\gamma t}$. Without loss of generality let S be the last $t - r$ coordinates:

$$\begin{array}{ll} X_1 \ldots X_r & X_{r+1} \ldots X_t \\ Y_1 \ldots Y_r & \underbrace{Y_{r+1} \ldots Y_t}_{\text{questions in } S} \end{array}$$

Write X_S, Y_S, A_S, B_S for the questions and answers in the coordinates of S. Recall that our goal is to find an i for which $\Pr[W_{\{i\}}|W_S]$ is small. In order to do this, we must use the fact that the value of the one shot game is bounded by $1 - \alpha$. Unfortunately, the bound of $1 - \alpha$ only holds when the questions are distributed according to X, Y. In our world, the questions are instead distributed according to $X_i, Y_i|W_S$. Thus, a first step towards proving the lemma is to show that there is some coordinate i in which the questions are distributed close to how they were before the conditioning.

This is not too hard to do.

$$\begin{aligned}
\gamma t &\geq D\left(p\left(\begin{matrix} x_1 \dots x_r \\ y_1 \dots y_r \end{matrix} \middle| W_S \right) \middle\| \, p\left(\begin{matrix} x_1 \dots x_r \\ y_1 \dots y_r \end{matrix} \right) \right) && \text{by Lemma 7} \\
&\geq \sum_{i=1}^{t} D\left(p\left(\begin{matrix} x_i \\ y_i \end{matrix} \middle| W_S \right) \middle\| \, p\left(\begin{matrix} x_i \\ y_i \end{matrix} \right) \right) && \text{by Lemma 10} \\
&\geq \sum_{i=1}^{t} \left| p\left(\begin{matrix} x_i \\ y_i \end{matrix} \middle| W_S \right) - p\left(\begin{matrix} x_i \\ y_i \end{matrix} \right) \right|^2 && \text{by Lemma 11} \\
\Rightarrow \gamma &\geq (1/t) \sum_{i=1}^{t} \left| p\left(\begin{matrix} x_i \\ y_i \end{matrix} \middle| W_S \right) - p\left(\begin{matrix} x_i \\ y_i \end{matrix} \right) \right|^2 \\
\Rightarrow \sqrt{\gamma} &\geq (1/t) \sum_{i=1}^{t} \left| p\left(\begin{matrix} x_i \\ y_i \end{matrix} \middle| W_S \right) - p\left(\begin{matrix} x_i \\ y_i \end{matrix} \right) \right| && \text{by convexity}
\end{aligned} \tag{1}$$

Thus, the distribution of the questions in the average coordinate changes only by $\sqrt{\gamma}$ in statistical distance.

Question: Why isn't this enough to prove the lemma?

One might hope that Alice and Bob can use the following strategy for the one shot game. Given a pair of questions X, Y, they use shared randomness to sample a uniformly random coordinate $i \in [t]$, set $X_i = x, Y_i = y$, jointly sample the rest of the questions conditioned on the event $W_S, X_i = x, Y_i = y$, and use the strategy of the repeated game to return the i'th answers A_i, B_i. This solution breaks down in the sampling step. Conditioning on W_S can create correlations between x, y and the rest of the questions. For example, nothing we have argued so far precludes the fact that $\Pr[Y_1 = X_i | W_S] = 1$. Thus, it could be that to correctly sample Y_1, Bob needs to know x, and since he receives no communication from Alice, it may be impossible for him to sample Y_1 correctly.

As in the proof of the lower bound for the communication complexity of set disjointness, and in the proof of the direct sum result for communication complexity, we are going to resolve this issue by breaking the dependence between Alice and Bob's questions.

Let $V = V_1, V_2, \dots, V_t$ be independent uniformly random bits. Define

$$T_i = \begin{cases} X_i & V_i = 0 \\ Y_i & \text{else} \end{cases}$$

$$U = \left(\begin{matrix} T_1, T_2, \dots, T_r, X_S \\ V_1, V_2, \dots, V_r, Y_S \end{matrix} \right)$$

We write U_{-i} to denote all of U excluding T_i, V_i.

Then observe that conditioned on any fixing of A_S, U, W_S, we have that $X^n | A_S, U, W_S$ is independent of $Y^n | A_S, U, W_S$. This is because U fixes at least one of the questions in every coordinate, A_S is determined by X^n, and once A_S, U are fixed, W_S is determined by Y^n. Let i be a uniformly random coordinate in $[r]$. We shall show that given questions x, y, Alice and Bob can use public randomness to correctly and consistently sample (with error at most $\varepsilon \leq \alpha - \gamma$) from the distribution $p(u, i, a_S | X_i = x, Y_i = y, W_S)$, without communicating. Then, they can then use private randomness to sample X^n, Y^n conditioned on the variables they have sampled, and then return the answers A_i, B_i. Since $\text{val}(\mathcal{C}) < 1-\alpha$, it must be the case that the probability of success of this strategy is less than $1-\alpha$. On the other hand, the probability of success is at least $\mathbb{E}_i\left[\Pr[W_i|W_S]\right] - \varepsilon$, proving that $\mathbb{E}_i\left[\Pr[W_i|W_S]\right] < 1-\gamma$ as required.

Next we show how Alice and Bob can sample from $p(u, i, a_S|X_i = x, Y_i = y, W_S)$. Recall that we use the notation $p(x) \overset{\varepsilon}{\approx} q(x)$ to denote the fact that $|p(x) - q(x)| \leq \varepsilon$. We shall prove the following two lemmas:

Lemma 12. *There is a choice for* $\gamma(\alpha, k)$ *so that*

$$p(i, x_i, y_i)\cdot p(u_{-i}, a_S|W_S, i, x_i) \overset{(\alpha-\gamma)/10}{\approx} p(i, u_{-i}, a_S, x_i, y_i|W_S) \overset{(\alpha-\gamma)/10}{\approx} p(i, x_i, y_i)\cdot p(u_{-i}, a_S|W_S, i, y_i).$$

Lemma 13. *There is a protocol for Alice and Bob to use public randomness to sample* x *in such a way that if Alice's input is* p *and Bob's input is* q*, Alice's sample is distributed according to* p*, Bob's sample is distributed according to* q*, and both parties obtain the same sample with probability at least* $1 - 2|p-q|$.

Sketch of Proof Let the public randomness consist of an infinite sequence of tuples

$$(x_1, \rho_1), (x_2, \rho_2), \ldots,$$

where each x_i is a uniformly random element of the universe, and ρ_i is a uniformly random real number in $[0, 1]$. Alice samples her element by picking the smallest i for which $p(x_i) \geq \rho_i$, and sampling x_i. Bob does the same using q. It is clear that both parties sample an element distributed according to their respective distributions. They make a mistake exactly when the ρ_i that they pick lies in between $p(x_i)$ and $q(x_i)$, which happens with probability bounded by $2|p-q|$. ■

Given Lemma 12, Alice and Bob can use public randomness to sample i, then set $x_i = x, y_i = y$ and use Lemma 13 to sample u_{-i}, a_S with, Alice using the distribution $p(u_{-i}, a_S|W_S, x_i)$ and Bob using the distribution $p(u_{-i}, a_S|W_S, y_i)$. This would complete the proof.

Proof of Lemma Lemma 12 Recall that $\Pr[W_S] \geq 2^{-\gamma t}$ and A_S has a support size of at most $k^{\gamma t}$. We apply Lemma 9 to conclude that

$$\gamma t + \gamma t \log k = \log(k^{\gamma t}/2^{\gamma t}) \geq \mathop{\mathbb{E}}_{U, A_S|W_S}\left[D(p(x^r, y^r|U, A, W_S)\|p(x^r, y^r|U)\right]$$

Since $X^r, Y^r|U$ consist of r independent coordinates, we can apply Lemma 10, to get the bound

$$\begin{aligned} \gamma t + \gamma t \log k &\geq \mathop{\mathbb{E}}_{U, A_S|W_S}\left[\sum_{j=1}^{r} D(p(x_j, y_j|U, A, W_S)\|p(x_j, y_j|U)\right] \\ \Rightarrow \gamma t(1 + \log k)/r &\geq \mathop{\mathbb{E}}_{i, U, A_S|W_S}\left[D(p(x_i, y_i|i, U, A, W_S)\|p(x_i, y_i|i, U)\right], \end{aligned}$$

where i is a uniformly random coordinate in $[r]$, independent of all other variables.

Next, apply Lemma 11 to conclude that

$$\begin{aligned} \gamma t(1 + \log k)/r &\geq \mathop{\mathbb{E}}_{i, U, A_S|W_S}\left[|p(x_i, y_i|i, U, A, W_S) - p(x_i, y_i|i, U)|^2\right] \\ &\geq \mathop{\mathbb{E}}_{i, U, A_S|W_S}\left[|p(x_i, y_i|i, U, A, W_S) - p(x_i, y_i|i, U)|\right]^2 \\ \Rightarrow \sqrt{\gamma t(1 + \log k)/r} &\geq \mathop{\mathbb{E}}_{i, U, A|W_S}\left[|p(x_i, y_i|U, A, W_S) - p(x_i, y_i|U)|\right] \end{aligned} \tag{2}$$

Set $\varepsilon = \sqrt{\gamma t(1+\log k)/r} = \sqrt{\gamma(1+\log k)/(1-\gamma)}$, and let γ be small enough so that $\varepsilon < (\alpha-\gamma)/100$.

Then Equation 2 implies that

$$\begin{aligned}
\varepsilon &\geq \mathop{\mathbb{E}}_{i,U,A_S|W_S}[|p(x_i,y_i|i,U,A,W_S) - p(x_i,y_i|i,U)|] \\
&= \mathop{\mathbb{E}}_{i,U_{-i},A_S|W_S}\left[\mathop{\mathbb{E}}_{U_i|i,A_S,W_S,U_{-i}}[|p(x_i,y_i|i,U,A,W_S) - p(x_i,y_i|i,U)|]\right] \\
&\geq \mathop{\mathbb{E}}_{i,U_{-i},A_S|W_S}\left[(1/2)\mathop{\mathbb{E}}_{X_i|i,A_S,W_S,U_{-i}}[|p(x_i,y_i|i,U_{-i},X_i,A,W_S) - p(x_i,y_i|i,U_{-i},X_i)|]\right], \qquad (3)
\end{aligned}$$

where the inequality follows by considering the part of the inner expectation that corresponds to $V_i = 0$.

Thus we have shown that

$$\begin{aligned}
p(i,x_i,y_i,u_{-i},a_S|W_S) &\overset{2\varepsilon}{\approx} p(i,x_i,u_{-i},a_S|W_S)\cdot p(y_i|i,u_{-i},x_i) \\
&= p(i,x_i,u_{-i},a_S|W_S)\cdot p(y_i|i,x_i) \\
&= p(i,x_i|W_S)\cdot p(y_i|i,x_i)\cdot p(u_{-i},a_S|i,x_i,W_S) \\
&\overset{\varepsilon}{\approx} p(i,x_i)\cdot p(y_i|i,x_i)\cdot p(u_{-i},a_S|i,x_i,W_S) \\
\Rightarrow p(i,x_i,y_i,u_{-i},a_S|W_S) &\overset{3\varepsilon}{\approx} p(i,x_i,y_i)\cdot p(u_{-i},a_S|i,x_i,W_S),
\end{aligned}$$

where in the last step, we used the fact that $p(i,x_i) \overset{\varepsilon}{\approx} p(i,x_i|W_S)$, which we proved in Equation 1.

Since $\varepsilon \leq (\alpha-\gamma)/100$, we obtain that

$$p(i,x_i,y_i,u_{-i},a_S|W_S) \overset{(\alpha-\gamma)/10}{\approx} p(i,x_i,y_i)\cdot p(u_{-i},a_S|i,x_i,W_S)$$

Similarly, we get

$$p(i,x_i,y_i,u_{-i},a_S|W_S) \overset{(\alpha-\gamma)/10}{\approx} p(i,x_i,y_i)\cdot p(u_{-i},a_S|i,y_i,W_S)$$

This concludes the proof of Lemma 12. ■

■

References

[1] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45, 1998.

[2] Hillel Furstenberg and Yitzhak Katznelson. A density version of the hales-jewett theorem. *Journal dAnalyse Mathematique*, 57, 1991.

[3] Thomas Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 2007.

[4] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998.

[5] Oleg Verbitsky. Towards the parallel repetition conjecture. In *Structure in Complexity Theory Conference*, pages 304–307, 1994.