THE PARALLEL REPETITION THEOREM

1. Introduction

The goal of this note is to give a complete proof of the *parallel repetition theorem* which is a fundamental result in theoretical computer science proved by Raz [4]. Our presentation follows essentially the approach from [3, 1] with some minor twists (in particular, we completely avoid the use of entropy).

A *game* $\mathscr{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$ is the data of

(1) finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}$
(2) π, a probability measure on $\mathcal{X} \times \mathcal{Y}$
(3) a function $V : \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B} \to \{0, 1\}$.

A *(deterministic) strategy* is a couple (f_A, f_B) of functions $f_A : \mathcal{A} \to \mathcal{X}$ and $f_B : \mathcal{B} \to \mathcal{Y}$. The *value* $\omega(\mathscr{G})$ of a game $\mathscr{G}$ is defined as

$$\omega(\mathscr{G}) := \sup_{(f_A, f_B) \text{ strategies}} \mathbf{P}(V(X, Y, f_A(X), f_B(Y)) = 1) \tag{1}$$

where (X, Y) is a random variable with distribution π. We describe the game using the concept of one referee and two players (named Alice and Bob): the referee select a pair $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ of random questions with distribution π, Alice answers $f_A(X)$ to the question X and Bob answers $f_B(Y)$ to the question Y.

Alternatively, we could consider *randomized strategies*, which are random variables taking values in the set of deterministic strategies. This gives the same value for the supremum. In this case it should be understood that (f_A, f_B) is independent from (X, Y) in (1).

Let $n \geqslant 1$. The n-parallel repetition of $\mathscr{G}$ is the game $\mathscr{G}^n$ defined as

$$\mathscr{G}^n := (\mathcal{X}^n, \mathcal{Y}^n, \mathcal{A}^n, \mathcal{B}^n, \pi^{\otimes n}, V_n)$$

where for $\bar{x} \in \mathcal{X}^n$, $\bar{y} \in \mathcal{Y}^n$, $\bar{a} \in \mathcal{A}^n$, $\bar{b} \in \mathcal{B}^n$,

$$V_n(\bar{x}, \bar{y}, \bar{a}, \bar{b}) := \prod_{i=1}^{n} V(x_i, y_i, a_i, b_i).$$

In other words, the players play n rounds of the game, where they are asked i.i.d. pairs of questions, and win the game $\mathscr{G}^n$ if they win each round of $\mathscr{G}$. We have

$$\omega(\mathscr{G})^n \leqslant \omega(\mathscr{G}^n) \leqslant \omega(\mathscr{G}). \tag{2}$$

Indeed, a possible strategy for $\mathscr{G}^n$ is ato nswer the jth round as a function of the jth question only; in that case the rounds are independent instances of $\mathscr{G}$, leading to the left inequality in (2). The right inequality follows by observing that in order to win $\mathscr{G}^n$, the players must win the first round. It is instructive to describe an example with $\omega(\mathscr{G}^2) = \omega(\mathscr{G}) \in (0, 1)$.

For this equality to happen, Alice and Bob must correlate their answers in such a way that they either win both rounds or lose both rounds.

Example 1. Consider the game $\mathscr{G}$ given by $\mathcal{X} = \{1,2\}$, $\mathcal{Y} = \{3,4\}$, $\mathcal{A} = \mathcal{B} = \{1,2,3,4\}$, π the uniform measure and V defined as

$$V(x,y,a,b) := \begin{cases} 1 & \text{if } a = b = x \text{ or } a = b = y \\ 0 & \text{otherwise} \end{cases}$$

One can check that $\omega(\mathscr{G}) = 1/2$. However, $\omega(\mathscr{G}^2) = 1/2$ as showed by the following strategy

$$f_A(x_1,x_2) := (x_1, x_1+2), \;\; f_B(y_1,y_2) = (y_2 - 2, y_2)$$

which wins whenever $x_1 + 2 = y_2$.

The *parallel repetition theorem* states that for any game $\mathscr{G}$ with $\omega(\mathscr{G}) = 1 - \delta < 1$, the quantity $\omega(\mathscr{G}^n)$ tends to 0 exponentially fast, at a rate which depends on δ and on $\Sigma = |\mathcal{A} \times \mathcal{B}|$.

Theorem 1 (Raz [4], Holenstein [3]). *If $\mathscr{G}$ is a game with $\omega(\mathscr{G}) = 1 - \delta$, then*

$$\omega(\mathscr{G}^n) \leqslant \exp\left(-\frac{c\delta^3 n}{\log \Sigma}\right)$$

where $c > 0$ is a constant.

2. The main lemma

We fix a game $\mathscr{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$, an integer n and a deterministic strategy (F_A, F_B) for $\mathscr{G}^n$. We consider an instance of the game $\mathscr{G}^n$ where players use the strategy (F_A, F_B). Let $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ be the questions (i.e., i.i.d. random variables with distribution π) and $(A_i, B_i)_{1 \leqslant i \leqslant n}$ be the answers defined as $(A_1, \dots, A_n) = F_A(X_1, \dots, X_n)$, $(B_1, \dots, B_n) = F_B(Y_1, \dots, Y_n)$. For $1 \leqslant i \leqslant n$, we consider the event

$$W_i := \{V(X_i, Y_i, A_i, B_i) = 1\}$$

that the ith round of the game is a win.

Here is the main lemma. Here $c > 0$ is an absolute constant.

Lemma 2 (Main lemma). *Assume that $k \leqslant \frac{c\delta^2 n}{\log \Sigma}$ and $\mathbf{P}(W_1 \cap \dots \cap W_k) \geqslant \Sigma^{-k}$. Then*

$$\frac{1}{n-k} \sum_{j=k+1}^{n} \mathbf{P}(W_j | W_1 \cap \dots \cap W_k) \leqslant 1 - \frac{\delta}{4}.$$

We show how Lemma 2 implies Theorem 1. Up to reordering the rounds of the games, we may assume that

$$\mathbf{P}(W_{k+1} | W_1 \cap \dots \cap W_k) \leqslant 1 - \frac{\delta}{4}$$

whenever $k \leqslant k_0 = \frac{c\delta^2 n}{\log \Sigma}$ and $\mathbf{P}(W_1 \cap \dots \cap W_k) \geqslant \Sigma^{-k}$. If we set $p_k = \mathbf{P}(W_1 \cap \dots \cap W_k)$, we have

$$p_k \leqslant \max\left(\Sigma^{-k+1}, (1 - \delta/4) p_{k-1}\right) \leqslant (1 - \delta/4)^k$$

and therefore,

$$p_n \leqslant p_{k_0} \leqslant (1-\delta/4)^{k_0} \leqslant \exp(-k_0\delta/4) = \exp(-c\delta^3 n/\log\Sigma)$$

and Theorem 1 follows by taking the supremum over strategies (F_A, F_B).

2.1. **Coupling, total variations, shared randomness.** Let μ_1, μ_2 be probability measures on a finite set S. The total variation distance between μ_1 and μ_2 is

$$\Delta(\mu_1 \ : \ \mu_2) \coloneqq \frac{1}{2}\sum_{x\in S}|\mu_1(x)-\mu_2(x)| = 1-\sum_{x\in S}\min(\mu_1(x),\mu_2(x)).$$

If X, Y are random variables with respective distributions μ, ν, we sometime write $\Delta(X \ : \ Y)$ instead of $\Delta(\mu \ : \ \nu)$. A basic coupling lemma states there is a probability space Ω and random variables X, Y defined on Ω, with respective laws μ, ν such that $\mathbf{P}(X \neq Y) = \Delta(\mu \ : \ \nu)$. Here is a more advanced version.

Lemma 3 (shared randomness)**.** *Let S be a finite set. There is a probability space Ω and, for every probability measure μ on S, a random variable $X_\mu : \Omega \to S$ with distribution μ such that, for every probability measures μ, ν on S,*

$$\mathbf{P}(X_\mu \neq X_\nu) \leqslant 2\Delta(\mu \ : \ \nu)$$

Proof. Denote by m the uniform measure on $S \times [0,1]$ (i.e., the product of the discrete uniform measure on S and of the continuous uniform measure on $[0,1]$). We associate to a probability measure μ on S its histogram $H_\mu \subset S \times [0,1]$ defined as

$$H_\mu \coloneqq \{(x,t) \in S \times [0,1] \ : \ t \leqslant \mu(x)\}.$$

Observe that $m(H_\mu) = 1/|S|$ and that $m(H_\mu \triangle H_\nu) = 2\Delta(\mu \ : \ \nu)/|S|$, where $\triangle$ denotes the symmetric difference. If $(Y_n)_n$ is a i.i.d. sequence of random variables with distribution m, we may define for every probability measure μ a random variable

$$X_\mu \coloneqq \inf\{n \ : \ Y_n \in H_\mu\}.$$

The distribution of X_μ is precisely μ. If ν is another probability measure, then

$$\mathbf{P}(X_\mu \neq X_\nu) \leqslant \frac{m(H_\mu \triangle H_\nu)}{m(H_\mu \cup H_\nu)}.$$

Indeed, the event $\{X_\mu = X_\nu\}$ is verified whenever the infima in the definition of X_μ and X_ν coincide, which happens whenever the first element of the sequence (Y_n) which belongs to $H_\mu \cup H_\nu$ actually belongs to $H_\mu \cap H_\nu$. The result follows since $m(H_\mu \cup H_\nu) \geqslant m(H_\mu) = 1/|S|$. $\square$

We use Lemma 3 in the following form: from shared randomness, Alice can generate $X \sim \mu$, Bob can generate $Y \sim \nu$ such that $\mathbf{P}(X \neq Y) \leqslant 2\Delta(\mu \ : \ \nu)$. This does not require Alice to know the measure ν, nor Bob to know the measure μ.

2.2. **A first look at the strategies.** We describe a randomized strategy (f_A, f_B) for $\mathscr{G}$ which depends on integers $k < n$ and on a deterministic strategy (F_A, F_B) for the game $\mathscr{G}^n$.

Here is a sketch of the strategy. Alice and Bob select an integer $j \in \{k+1, \dots, n\}$. When Alice is asked a question x, she generates a random n-tuple of questions

$$\overline{\xi} = (\xi_1, \dots, \xi_n) \in \mathcal{X}^n$$

with $\xi_j = x$, and defines $f_A(x)$ as the jth coordinate of $F_A(\xi)$. Similarly, when Bob is asked a question y, he generates a random n-tuple questions

$$\overline{\eta} = (\eta_1, \dots, \eta_n) \in \mathcal{Y}^n$$

with $\eta_j = y$, and defines $f_B(y)$ as the jth coordinate of $F_B(\eta)$.

In order for this strategy to be efficient, Alice and Bob need to correlate their randomness generation. How they achieve this is explained in the next sections.

2.3. **Clues.** We introduce a more complicated equivalent version of the game $\mathscr{G}^n$. In this version, in a first stage, the referee reveals a random selection of half of the questions (which we call the clues) and reveals the full list of questions only in a second stage.

Introduce a symbol $\star$ which is distinct from elements in $\mathcal{X}$ and from elements in $\mathcal{Y}$. The symbol $\star$ will play the role of an unknown element. Define the set of *clues* to be

$$\mathcal{C} \coloneqq (\mathcal{X} \times \{\star\}) \cup (\{\star\} \times \mathcal{Y}).$$

A clue is a pair of questions, one of them being unknown. Let (X, Y) be a pair of questions with distribution π, and C be the random clue defined as either $(X, \star)$ of $(\star, Y)$ with probability $1/2$. We denote bt $\hat{\pi}$ the distribution of C. For $c \in \mathcal{C}$, let π^1_c the distribution of $X|[C = c]$ (when $c = (x, \star)$, this distribution is the Dirac mass δ_x; when $c = (\star, y)$, this distribution is proportional to $\pi(\cdot, y)$). Similarly, let π^2_c the distribution of $Y|[C = c]$.

An equivalent way to generate the questions $(X_i, Y_i)_{1 \leqslant i \leqslant n}$ is as follows

1. Generate i.i.d. clues $(C_i)_{1 \leqslant i \leqslant n}$ with distribution $\hat{\pi}$,
2. For each $1 \leqslant i \leqslant n$, generate X_i according to the distribution $\pi^1_{C_i}$ and Y_i according to the distribution $\pi^2_{C_i}$. All choices are assumed to be independent (note that half of the $2n$ variables (X_i, Y_i) are in fact deterministic given (C_i).)

2.4. **A non-admissible strategy.** We define a *transcript* to be an element of $\mathcal{T} \coloneqq \mathcal{X}^k \times \mathcal{Y}^k \times \mathcal{A}^k \times \mathcal{B}^k$.

We consider an instance of the game $\mathscr{G}^n$ where the questions are revealed in the 2-step procedure via clues. Let (C_i) be the clues, (X_i, Y_i) be the questions and (A_i, B_i) be the answers. The transcript of the game is the random variable

$$T \coloneqq (X_i, Y_i, A_i, B_i)_{1 \leqslant i \leqslant k} \in \mathcal{T}. \tag{3}$$

For $j \in \{k + 1, \dots, n\}$, we denote by $C_{\neg j}$ the random variable

$$C_{\neg j} \coloneqq (C_{k+1}, \dots, C_{j-1}, C_{j+1}, \dots, C_n) \in \mathcal{C}^{n-k-1}.$$

We now explain how Alice and Bob generate the questions $\overline{\xi}$ and $\overline{\eta}$ in a correlated way. We first describe a strategy which is not acceptable since it requires communication between Alice and Bob at the stage (3). We then modify the strategy to remove communication.

(1) The strategy depends on an index $j \in \{k+1, \dots, n\}$. Alice sets $\xi_j = x$ and Bob sets $\eta_j = y$.
(2) The strategy depends on a transcript $t = (\overline{x}, \overline{y}, \overline{a}, \overline{b}) \in \mathcal{T}$. Alice sets $(\xi_1, \dots, \xi_k) = \overline{x}$ and Bob sets $(\eta_1, \dots, \eta_k) = \overline{y}$.
(3) Using shared randomness, Alice and Bob generate a random list of clues C with distribution $C_{\neg j}|[T = t, X_j = x, Y_j = y]$. Set $C_A = C_B = C$.
(4a) If $C_A = (c_{k+1}, \dots, c_{j-1}, c_{j+1}, \dots, c_n)$, then Alice generates using local randomness $(\xi_{k+1}, \dots, \xi_{j-1}, \xi_{j+1}, \dots, \xi_n)$ according to the distribution $\pi^1_{c_{k+1}} \otimes \cdots \otimes \pi^1_{c_{j-1}} \otimes \pi^1_{c_{j+1}} \otimes \cdots \otimes \pi^1_{c_n}$ conditioned to the event $(F_A(\overline{\xi})_i)_{1 \leqslant i \leqslant k} = \overline{a}$.
(4b) If $C_B = (c_{k+1}, \dots, c_{j-1}, c_{j+1}, \dots, c_n)$, then Bob generates using local randomness $(\eta_{k+1}, \dots, \eta_{j-1}, \eta_{j+1}, \dots, \eta_n)$ according to the distribution $\pi^2_{c_{k+1}} \otimes \cdots \otimes \pi^2_{c_{j-1}} \otimes \pi^2_{c_{j+1}} \otimes \cdots \otimes \pi^2_{c_n}$ conditioned to the event $(F_B(\overline{\eta})_i)_{1 \leqslant i \leqslant k} = \overline{b}$.
(5) Alice defines $f_A(x)$ as the jth coordinate of $F_A(\xi)$, and Bob defines $f_B(x)$ as the jth coordinate of $F_B(\eta)$

In this procedure, the random variables $(\overline{\xi}, \overline{\eta}$ have the same distribution as the random variables $(X_i)_{1 \leqslant i \leqslant n}, (Y_i)_{1 \leqslant i \leqslant n}$ conditioned to $W_1 \cap \cdots \cap W_k \cap \{(X_j, Y_j) = (x, y)\}$. This strategy loses the game $\mathscr{G}$ with probability

$$\mathbf{P}(\overline{W_j} | T = t, X_j = x, Y_j = y)$$

2.5. **Correcting the strategy.** We modify the strategy described at the previous section, to eliminate communication between Alice and Bob. We replace step (3) by the following

(3a) Alice generates a random list of clues C_A with distribution $C_{\neg j}|[T = t, X_j = x]$.
(3b) Bob generates a random list of clues C_B with distribution $C_{\neg j}|[T = t, Y_j = y]$.

This step must be done in a correlated way using shared randomness. By Lemma 3, there exists a probability space Ω and for every $x \in \mathcal{X}, y \in \mathcal{Y}$ random variables $\Gamma_x, \Gamma^y, \Gamma^y_x : \Omega \to \mathcal{C}^{n-k-1}$ such that

- the random variable Γ_x has the distribution of $C_{\neg j}|[T = t, X_j = x]$
- the random variable Γ^y has the distribution of $C_{\neg j}|[T = t, Y_j = y]$
- the random variable Γ^y_x has the distribution of $C_{\neg j}|[T = t, X_j = x, Y_j = y]$
- $\mathbf{P}(\Gamma_x = \Gamma^y = \Gamma^y_x) \geqslant 1 - \delta(t, j, x, y)$

where

$$\begin{aligned}\delta(t, j, x, y) \coloneqq{} & 2\Delta(C_{\neg j}|[T = t, X_j = x] \ : \ C_{\neg j}|[T = t, X_j = x, Y_j = y]) \\ & + 2\Delta(C_{\neg j}|[T = t, Y_j = y] \ : \ C_{\neg j}|[T = t, X_j = x, Y_j = y])\end{aligned}$$

In order to correlate their random selection, we require that Alice and Bob perform steps (3a) and (3b) by setting $C_A = \Gamma_x$ and $C_B = \Gamma^y$. Note that setting $C_A = C_B = \Gamma^y_x$ would exactly implement step (3). It follows that this strategy, when asked (x, y), loses with probability at most

$$\ell(x,y) \coloneqq \mathbf{P}(\overline{W_j}|T=t, X_j=x, Y_j=y) + \delta(t,j,x,y)$$

When asked a random question with distribution π, this strategy loses with probability at most

$$\begin{aligned} \sum_{x,y} \pi_{x,y}\ell(x,y) &\leqslant \sum_{x,y} \mathbf{P}(X_j=x, Y_j=y|T=t)\ell(x,y) + \Delta_1(t,j) \\ &\leqslant \mathbf{P}(\overline{W_j}|T=t) + \Delta_1(t,j) + \Delta_2(t,j) \end{aligned}$$

where $\Delta_1(t,j) \coloneqq \Delta(\pi \ : \ (X_j,Y_j)|[T=t])$ and

$$\begin{aligned} \Delta_2(t,j) &\coloneqq \sum_{x,y} \mathbf{P}(X_j=x, Y_j=y|T=t)\delta(t,j,x,y) \\ &= 2\Delta((X_j,Y_j)|[T=t], C_{\neg j}|[T=t,X_j] \ : \ (X_j,Y_j,C_{\neg j})|[T=t]) \\ &\quad +2\Delta((X_j,Y_j)|[T=t], C_{\neg j}|[T=t,X_j] \ : \ (X_j,Y_j,C_{\neg j})|[T=t]). \end{aligned}$$

By definition of $\omega(\mathscr{G})$, this strategy loses the game $\mathscr{G}$ with probability at least δ. Therefore the inequality

$$\delta \leqslant \mathbf{P}(\overline{W_j}|T=t) + \Delta_1(t,j) + \Delta_2(t,j) \tag{4}$$

holds for every index j and every transcript t.

We now state a fundamental lemma which bounds the parameters $\Delta_1(t,j)$ and $\Delta_2(t,j)$ which appeared as the error when we modified the strategy to forbid communication. Given a transcript $t=(\overline{x},\overline{y},\overline{a},\overline{b}) \in \mathcal{T}$, let $p(t)$ be the probability that the transcript t occurs when $\mathscr{G}^n$ is played with strategy (F_A,F_B), where the first k rounds of questions are deterministic according to $(\overline{x},\overline{y})$, and the next rounds are independent with distribution π.

Lemma 4. *Let t be a transcript. Then*

$$\sum_{j=k+1}^{n} \Delta_1(t,j) \leqslant 2\sqrt{n\log(1/p(t))}$$

$$\sum_{j=k+1}^{n} \Delta_2(t,j) \leqslant 8\sqrt{n\log(1/p(t))}$$

2.6. **Proof of the main Lemma.** We complete the proof of the main Lemma (Lemma 2) assuming Lemma 4. Let $t=(\overline{x},\overline{y},\overline{a},\overline{b}) \in \mathcal{X}^k \times \mathcal{Y}^k \times \mathcal{A}^k \times \mathcal{B}^k$ a transcript. We say that t is *likely* if $p(t) \geqslant \Sigma^{-3k}$, and that t is *winning* if $V(x_i,y_i,a_i,b_i)=1$ for every $1 \leqslant i \leqslant k$.

Consider a game where questions $(X_i,Y_i)_{1\leqslant i\leqslant n}$ are i.i.d. with distribution π and where the players use the strategy (F_A,F_B). Let T be the random transcript given by (3). For every $\overline{x},\overline{y} \in \mathcal{X}^k \times \mathcal{Y}^k$, by the union bound when summing over the Σ^k possible $\overline{a},\overline{b}$,

$$\mathbf{P}(T \text{ unlikely}|(X_1,\ldots,X_k)=\overline{x},(Y_1,\ldots,Y_k)=\overline{y}) \leqslant \Sigma^k\Sigma^{-3k} = \Sigma^{-2k}$$

and therefore the same bounds holds without conditioning on $\overline{x}, \overline{y}$. It follows that

$$\mathbf{P}(T \text{ unlikely}|W_1 \cap \cdots \cap W_k) \leqslant \frac{\mathbf{P}(T \text{ unlikely})}{\mathbf{P}(W_1 \cap \cdots \cap W_k)} \leqslant \frac{\Sigma^{-2k}}{\Sigma^{-k}} = \Sigma^{-k} \leqslant 1/2$$

Whenever t is a likely transcript, we have $\log(1/p(t)) \leqslant 3k \log \Sigma$ and therefore

$$\frac{1}{n-k} \sum_{j=k+1}^{n} \Delta_1(t,j) + \Delta_2(t,j) \leqslant \frac{10\sqrt{3nk \log \Sigma}}{n-k} \leqslant \frac{\delta}{2}$$

if we assume that $k \leqslant c\delta^2 n / \log \Sigma$ for a well-chosen $c > 0$. Finally, using (4),

$$\begin{aligned}
&\frac{1}{n-k} \sum_{j=k+1}^{n} \mathbf{P}(\overline{W_j}|W_1 \cap \cdots \cap W_k) \\
= \quad & \sum_{t \in \mathcal{T}} \mathbf{P}(T = t|W_1 \cap \cdots \cap W_k) \frac{1}{n-k} \sum_{j=k+1}^{n} \mathbf{P}(\overline{W_j}|T = t) \\
\geqslant \quad & \sum_{t \text{ likely}} \mathbf{P}(T = t|W_1 \cap \cdots \cap W_k) \frac{1}{n-k} \sum_{j=k+1}^{n} (\delta - \Delta_1(t,j) - \Delta_2(t,j)) \\
\geqslant \quad & \sum_{t \text{ likely}} \mathbf{P}(T = t|W_1 \cap \cdots \cap W_k) \frac{\delta}{2} \\
= \quad & \frac{\delta}{2} \mathbf{P}(T \text{ likely}|W_1 \cap \cdots \cap W_k) \\
\geqslant \quad & \frac{\delta}{4}
\end{aligned}$$

and Lemma 2 follows.

3. Proof of Lemma 4

The following lemma appears as [3, Lemma 5]. We give a different and arguably simpler proof in Section 4, by using Hoeffding's inequality instead of considerations about entropy (our works in the case, say, $\mathbf{P}(E) < 1/10$, which is the range in which we apply it).

Lemma 5. *Let $Z_1, \ldots, Z_n$ be independent random variables and E an event. Then*

$$\sum_{j=1}^{n} \Delta(Z_j|E \ : \ Z_j)^2 \leqslant \log(1/\mathbf{P}(E))$$

Lemma 6. *Let S be a random variable, $Z_1, \ldots, Z_n$ be random variables conditionally independent given S, and E an event. Then*

$$\sum_{j=1}^{n} \Delta(\mu_j \ : \ \nu_j) \leqslant \sqrt{n}\sqrt{\log(1/\mathbf{P}(E))},$$

where μ_j is the distribution of $(SZ_j)|E$ and ν_j is the distribution of a pair (s, z_j), where s is distributed as $S|E$ and z_j is distributed as $Z_j|[S = s]$.

Let us show how Lemmas 5 and 6 imply Lemma 4. Fix a transcript $t = (\overline{x}, \overline{y}, \overline{a}, \overline{b}) \in \mathcal{T}$. Consider independent random variables $Z_j = (X_j, Y_j)$, where Z_j is deterministic and equal to (x_j, y_j) for $j \leqslant k$ and of distribution π for $j > k$, and let T the corresponding random transcript. Consider the event $E = \{T = t\}$. Observe that $p(t)$ coincides with $\mathbf{P}(E)$. For $k < j \leqslant n$, we have

$$\Delta_1(t, j) = \Delta(Z_j \ : \ Z_j|E)$$

and therefore

$$\sum_{j=k+1}^{n} \Delta_1(t, j) \leqslant \sqrt{n} \left(\sum_{j=k+1}^{n} \Delta(Z_j|E \ : \ \pi)^2 \right)^{1/2} \leqslant \sqrt{n \log(1/p(t))}$$

and the result follows.

For the second part of Lemma 4, consider $S = (C_j)_{k<j\leqslant n}$, the clues in the 2-step procedure to generate the questions (X_j, Y_j). It is indeed the case that (Z_j) are conditionally independent given S. We apply Lemma 6 to obtain

$$\sum_{j=k+1}^{n} \Delta(\mu_j \ : \ \nu_j) \leqslant \sqrt{n}\sqrt{\log(1/p(T))}$$

where μ_j is the distribution of $(S, (X_j, Y_j))$ and ν_j is the distribution of $(S, (X_j, Y_j)|S)$. By reasoning on whether $C_j = (X_j, \star)$ or $C_j = (\star, Y_j)$, we obtain

$$\Delta(\mu_j \ : \ \nu_j) = \frac{1}{2}\Delta(C_{\neg j}, X_j, Y_j \ : \ C_{\neg j}, X_j, Y_j|X_j) + \frac{1}{2}\Delta(C_{\neg j}, X_j, Y_j \ : \ C_{\neg j}, X_j|Y_j, Y_j).$$

This bound implies

$$\Delta(C_{\neg j}, X_j, Y_j|X_j \ : \ C_{\neg j}, X_j|Y_j, Y_j) \leqslant 4\Delta(\mu_j \ : \ \nu_j).$$

Finally, observe that

$$\Delta_2(T, j) := \Delta(C_{\neg j}|X_j, X_j, Y_j \ : \ C_{\neg j}|Y_j, X_j, Y_j) = \Delta(C_{\neg j}, X_j, Y_j|X_j \ : \ C_{\neg j}, X_j|Y_j, Y_j)$$

Therefore, we have

$$\sum_{j=k+1}^{n} \Delta_2(T, j) \leqslant 4 \sum_{j=k+1}^{n} \Delta(\mu_j \ : \ \nu_j) \leqslant 4\sqrt{n}\sqrt{\log(1/p(T))}$$

as needed.

4. Proof of Lemmas 5 and 6

Lemma 7 (a la Hoeffding)**.** *Let (X_i) be independent random variables, with X_i taking values in an interval of length θ_i. Assume $\sum \theta_i^2 = 1$, and set $S := \sum X_i$. Then for any event E with $\mathbf{P}(E) < 1/10$,*

$$\Big| \mathbf{E}[S|E] - \mathbf{E}[S] \Big| \leqslant \sqrt{\log_2(1/\mathbf{P}(E))}$$

Proof. We recall Hoeffding's inequality [2]: under the same assumption, for any $t \geqslant 0$

$$\mathbf{P}\left(S \geqslant \mathbf{E}[S] + t\right) \leqslant \exp\left(-2t^2\right).$$

We may assume that $\mathbf{E}[S] = 0$. Assume also that $\mathbf{P}(E) \leqslant 1/2$. Write, for β to be determined.

$$\mathbf{E}[S|E] \leqslant \beta + \int_\beta^\infty \mathbf{P}(S \geqslant t|E)\,\mathrm{d}t \leqslant \beta + \int_\beta^\infty \frac{\mathbf{P}(S \geqslant t)}{\mathbf{P}(E)}\,\mathrm{d}t \leqslant \beta + \frac{1}{\mathbf{P}(E)}\int_\beta^\infty \exp(-2t^2)\,\mathrm{d}t.$$

Using the bound $\int_\beta^\infty \exp(-2t^2)\,\mathrm{d}t \leqslant \int_\beta^\infty \frac{t}{\beta}\exp(-2t^2)\,\mathrm{d}t = \exp(-2\beta^2)/4\beta$, we obtain

$$\mathbf{E}[S|E] \leqslant \beta + \frac{\exp(-2\beta^2)}{4\beta\mathbf{P}(E)}$$

and the choice $\beta = \sqrt{\ln(1/\mathbf{P}(E))/2}$ gives

$$\mathbf{E}[S|E] \leqslant \frac{\sqrt{\ln(1/\mathbf{P}(E))}}{\sqrt{2}} + \frac{\sqrt{2}}{4\sqrt{\ln(1/\mathbf{P}(E))}} \leqslant \sqrt{\ln(1/\mathbf{P}(E))}\left[\frac{1}{\sqrt{2}} + \frac{\sqrt{2}}{4\ln 10}\right]$$

using that $\mathbf{P}(E) \leqslant 1/10$. Since $\frac{1}{\sqrt{2}} + \frac{\sqrt{2}}{4\ln 10} \leqslant \frac{1}{\sqrt{\ln 2}}$, Lemma 7 follows by applying the same inequality to $-S$. $\square$

We now prove Lemma 5. We need to show that

$$\left(\sum_{j=1}^n \Delta(Z_j|E \ :\ Z_j)^2\right)^{1/2} \leqslant \sqrt{\log_2(1/\mathbf{P}(E))}. \tag{5}$$

First, note that (5) is equivalent to the fact that for every $\theta \in S^{n-1}$

$$\sum_{j=1}^n \theta_j \Delta(Z_j|E \ :\ Z_j) \leqslant \sqrt{\log_2(1/\mathbf{P}(E))}.$$

Next, for any random variables U, V, we have $\Delta(U\ :\ V) = \sup \mathbf{E}[f(U) - f(V)]$ where the supremum is over functions $f : \mathbf{R} \to \{-1/2, 1/2\}$. It suffices therefore to show that for every functions $f_j : \mathbf{R} \to \{-1/2, 1/2\}$

$$\sum_{j=1}^n \theta_j\left(\mathbf{E}[f_j(Z_j)|E] - \mathbf{E}[f_j(Z_j)]\right) \leqslant \sqrt{\log_2(1/\mathbf{P}(E))},$$

which follows from Lemma 7.

We now turn to the proof of Lemma 6. We write, using successively the inequality between the ℓ_1 and ℓ_2 norms on $\mathbf{R}^n$, Lemma 5 and the concavity of $x \mapsto \sqrt{\log x}$ for $x \geqslant 1$

$$
\begin{aligned}
\sum_{j=1}^{n} \Delta(\mu_j \ : \ \nu_j) &= \sum_{s \in S} \mathbf{P}(S=s|E) \sum_{j=1}^{n} \Delta(Z_j|E, S=s \ : \ Z_j|S=s) \\
&\leqslant \sqrt{n} \sum_{s \in S} \mathbf{P}(S=s|E) \left(\sum_{j=1}^{n} \Delta(Z_j|E, S=s \ : \ Z_j|S=s)^2 \right)^{1/2} \\
&\leqslant \sqrt{n} \sum_{s \in S} \mathbf{P}(S=s|E) \sqrt{\log(1/\mathbf{P}(E|S=s))} \\
&\leqslant \sqrt{n} \sqrt{\log\left(\sum_{s \in S} \frac{\mathbf{P}(S=s|E)}{\mathbf{P}(E|S=s)} \right)} \\
&= \sqrt{n} \sqrt{\log\left(\sum_{s \in S} \frac{\mathbf{P}(S=s)}{\mathbf{P}(E)} \right)} \\
&= \sqrt{n} \sqrt{\log(1/\mathbf{P}(E))}
\end{aligned}
$$

References

[1] Boaz Barak. Parallel repetition lemma. `https://www.cs.princeton.edu/courses/archive/spring07/cos522/ho11.pdf`.

[2] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[3] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. *Theory Comput.*, 5:141–172, 2009.

[4] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.