Theorem 5.1.6 (Parallel Repitition)**.** $\forall s \in (0,1),\ \exists c_s \in (0,s),\ PCP^{\Sigma}_{1,s}[r,2] \subseteq PCP^{\Sigma^t}_{1,c_s^t}[rt,2]$.

Corollary 5.1.7. *$\forall \varepsilon > 0,\ \exists$ an alphabet Σ such that*

$$NP \subseteq PCP^{\Sigma}_{1,\varepsilon}[O(\log n \cdot \log(1/\varepsilon)),2]$$

where such that $|\Sigma| = \mathrm{poly}(1/\varepsilon)$.

5.1.1 LABEL-COVER

We saw that there are PCP proof systems for languages in NP, in which the verifier makes only 2 queries. Furthermore, observe that the check made by the verifier is in the form of a projection from the first answer to the second. It will be more convenient to abstract these 2-query PCPs in terms of a graph problem, which we call LABEL-COVER.

Definition 5.1.8 (LABEL-COVER)**.** *An instance I of the* LABEL-COVER *problem is specified by a quadruple (G,Σ_1,Σ_2,F) where $G = (L,R,E)$ is a bipartite graph, Σ_1 and Σ_2 are two finite sized alphabets and $\Pi = \{\pi_e : \Sigma_1 \to \Sigma_2 | e \in E\}$, is a set of functions (also called projections), one for each edge $(u,v) \in E$.*

A labeling $A : L \to \Sigma_1, B : R \to \Sigma_2$, is said to satisfy an edge (u,v) iff $\pi_{(u,v)}(A(u)) = B(v)$. The value *of an instance is the maximal fraction of edges satisfied by any such labeling.*

For any $\delta \in (0,1)$, the gap problem GAP$_\varepsilon$-LC *is the promise problem of deciding if a given instance has value 1 or at most ε. More precisely, the* YES *and* NO *of* GAP$_\varepsilon$-LC *are given as follows.*

$$\begin{aligned}
\mathsf{YES} &= \left\{ I : \exists\,(A : L \to \Sigma_1,\ B : R \to \Sigma_2)\ \text{such that}\ \forall (u,v) \in E,\ \pi_{(u,v)}(A(u)) = B(v) \right\} \\
\mathsf{NO} &= \left\{ I : \forall\,(A : L \to \Sigma_1,\ B : R \to \Sigma_2)\,,\ \left|\{(u,v) \in E : \pi_{(u,v)}(A(u)) =, B(v)\}\right| \le \varepsilon |E| \right\}
\end{aligned}$$

Thus, an equivalent formulation of Corollary 5.1.7 is the following.

Corollary 5.1.9. *$\forall \varepsilon > 0$, there exist alphabets Σ_1, Σ_2 with $|\Sigma_1|, |\Sigma_2| = \mathrm{poly}(1/\varepsilon)$ such that* GAP$_\varepsilon$-LC *is NP-hard.*

5.2 Linearity Testing

We will now see a randomized test for checking whether a function between two groups is linear (ie a homomorphism).

Definition 5.2.1 (Linear function)**.** *Gven two Abelian groups G, H, the function $f : G \to H$ is said to be linear iff*

$$\forall x, y \in G,\ \ f(x+y) = f(x) + f(y).$$

(Observe that the first + is performed according to group G while the second is performed according to H.

A naive test for checking for linearity is to randomly pick $x, y \in G$ and check if $f(x+y) = f(x)+f(y)$. Blum, Luby and Rubinfield [BLR93] showed that this check is actually a "good" one.

$$\begin{array}{rl} \mathsf{BLR\text{-}Test}^f : \text{``} & \text{1. Choose } y, z \in_R G \text{ independently} \\ & \text{2. Query } f(y), f(z), \text{ and } f(y+z) \\ & \text{3. Accept if } f(y) + f(z) = f(y+z). \qquad\qquad \text{''} \end{array}$$

We will prove that this test is "good". Before stating the result we need some definitions.

Definition 5.2.2 (Local consistency)**.** *Local consistency of $f \in H^G$ denoted by $\varepsilon(f)$ is defined as*

$$\varepsilon(f) = \Pr_{x,y}[f(x) + f(y) \neq f(x+y)]$$

It is clear that, if f is linear then the naive check is true with probability 1 (ie. $\varepsilon(f) = 0$).

Definition 5.2.3 (Hamming distance)**.** *For $f, g \in H^G$, the* Hamming distance *$\delta(f, g)$ is defined as*

$$\delta(f, g) = \Pr_{x \in G}[f(x) \neq g(x)]$$

For $f \in H^G$ and $S \subseteq H^G$,

$$\delta(f, S) = \min_{g \in S} \delta(f, g)$$

Definition 5.2.4 (Global Consistency)**.** *Let $L \subset H^G$ be the set of all linear functions (homomorphisms) from G to H. Then the* global consistency *of $f \in H^G$ denoted by $\delta(f)$ is defined as*

$$\delta(f) = \delta(f, L)$$

It is clear that $\delta(f) = 0 \Rightarrow \varepsilon(f) = 0$. To show that the test is "good" is equivalent to proving a result of the following type: $\delta(f)$ is "large" implies $\varepsilon(f)$ is "large". We will first show that a general result of this nature is impossible if $\varepsilon(f)$ is not too small.

Consider the function $f : \mathbb{Z}/3n\mathbb{Z} \to \mathbb{Z}/3n\mathbb{Z}$ defined as follows:

$$f(x) = \begin{cases} 0 & \text{if } x \equiv 0 \pmod 3 \\ 1 & \text{if } x \equiv 1 \pmod 3 \\ 3n-1 & \text{if } x \equiv -1 \pmod 3 \end{cases}$$

It can be shown that though for this f, $\varepsilon(f) = 2/9$, $\delta(f) = 2/3$. Thus, even though $\delta(f)$ is very large, $\varepsilon(f)$ is not all that large. This counterexample was given by Coppersmith. We will now give an analysis of the test (also due to Coppersmith) which shows that this is basically the worst example. and also showed the following

Claim 5.2.5. *Suppose $\varepsilon(f) < 2/9$ then $\delta(f) \leq 2\varepsilon(f)$.*

Proof. Let $\varphi : G \to H$ be defined as

$$\varphi(x) = \text{plurality}_y\{f(x+y) - f(y)\},$$

with ties being broken arbibtrarily. We will show that φ has the following properties, which clearly implies the claim.

1. $\delta(f, \varphi) \le 2\varepsilon(f)$
2. $\forall x, \Pr_y[\varphi(x) = f(x+y) - f(y)] \ge 2/3$. Thus, even though $\varphi(x)$ was defined as the plurality, it is actually a 2/3-majority.
3. φ is linear

1. Proof of "$\delta(f, \varphi) \le 2\varepsilon(f)$"

   Let BAD $= \{x \in G : \Pr_y[f(x) \ne f(x+y) - f(y)] \ge 1/2\}$. If $x \notin$ BAD, then $f(x) = \varphi(x)$. Hence, $\delta(f, \varphi) \le |\text{BAD}|/|G|$. Now

$$\begin{aligned} \varepsilon(f) &= \Pr_{x,y}[f(x) \ne f(x+y) - f(y)] \\ &\ge \Pr\left[x \in \text{BAD}\right] \cdot \Pr\left[f(x) \ne f(x+y) - f(y) | x \in \text{BAD}\right] \\ &\ge \frac{|\text{BAD}|}{2|G|} \end{aligned}$$

2. Proof of "$\forall x,\ \Pr_y[\varphi(x) = f(x+y) - f(y)] \ge 2/3$"

   Fix any x, consider the following the collision probability

$$\begin{aligned} & \Pr_{y_1,y_2}\left[f(x+y_1) - f(y_1) = f(x+y_2) - f(y_2)\right] \\ = & \Pr_{y_1,y_2}\left[f(x+y_1) + f(y_2) = f(x+y_2) + f(y_1)\right] \\ \ge & \Pr_{y_1,y_2}\left[f(x+y_1) + f(y_2) = f(x+y_1+y_2) = f(x+y_2) + f(y_1)\right] \\ \ge & \ 1 - 2\varepsilon(f) > 5/9 \end{aligned}$$

   For $h \in H$, let $p_h = \Pr_y[f(x+y) - f(y) = h]$. Clearly, $p_{\max} = \max p_h = \Pr_y[\varphi(x) = f(x+y) - f(y)]$. Since the p_h's are a probability distribution, we have $\sum_{h \in H} p_h = 1$. From the above argument wrt. to the collision probability we have $\sum_{h \in H} p_h^2 > 5/9$. Then the following is true

$$\begin{aligned} p_{\max} = p_{\max} \cdot \sum p_h &\ge \sum_{h \in H} p_h^2 > 5/9 \\ p_{\max}^2 + (1 - p_{\max})^2 &\ge \sum_{h \in H} p_h^2 > 5/9 \\ 2p_{\max}^2 - 2p_{\max} + 4/9 &> 0 \\ p_{max} &> 2/3 \end{aligned}$$

   So for at least 2/3 fraction of the y's $f(x+y) - f(y)$ is the same and hence equal to $\varphi(x)$.

3. φ is linear.
   From 2, we have that

$$\begin{aligned}
\varphi(x) &= f(y) - f(y-x) \text{ for all but } < 1/3 \text{ fraction of } y\text{'s} \\
\varphi(z) &= f(y+z) - f(y) \text{ for all but } < 1/3 \text{ fraction of } y\text{'s} \\
\varphi(x+z) &= f(y+z) - f(y-x) \text{ for all but } < 1/3 \text{ fraction of } y\text{'s}
\end{aligned}$$

   Therefore $\exists y$ such that all the above 3 equations are satisfied. Hence

$$\forall x, z \in H, \ \varphi(x) + \varphi(z) = \varphi(x+z)$$

□

References

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. *Self-testing/correcting with applications to numerical problems.* J. Computer and System Sciences, 47(3):549–595, December 1993. (Preliminary Version in *22nd STOC*, 1990). doi:10.1016/0022-0000(93)90044-W.

[FK00] Uriel Feige and Joe Kilian. *Two-prover protocols – low error at affordable rates.* SIAM J. Computing, 30(1):324–346, 2000. (Preliminary Version in 26*th STOC*, 1994). doi:10.1137/S0097539797325375.

[Raz98] Ran Raz. *A parallel repetition theorem.* SIAM J. Computing, 27(3):763–803, June 1998. (Preliminary Version in *27th STOC*, 1995). doi:10.1137/S0097539795280895.