1. In this exercise problem, we will construct an efficient PCP for the class NP with $q$ queries. Suppose we want to verify a proof written in binary using $q$ queries. In this setting, a random ‘proof’ (and hence a proof of a wrong statement) is accepted with probability at least $2^{-q}$, assuming non trivial local views. Thus, with each extra query, we cannot hope to reduce the soundness by a (multiplicative) factor less than $\frac{1}{2}$.

We can quantify this parameter by looking at the ratio $\bar{q} := \frac{q}{\log_2(1/s)}$ where $s$ is the acceptance probability of incorrect proofs (i.e soundness of the PCP). This ratio is called the amortized query complexity of the PCP (i.e the number of queries needed to reduce the soundness by $\frac{1}{2}$ on average). The main goal of this exercise is to construct a PCP with $1 + o(1)$ amortized query complexity.

(a) As we have seen in the proof of NP-hardness of gap-3LIN($\frac{1}{2} - \epsilon, 1 - \epsilon$), one can replace BLR linearity test with any other $q$-query linearity test for some constant $q$. Consider replacing the BLR linearity test with the following test for checking a given function $f : \{0,1\}^n \rightarrow \{0,1\}$

$t$-fold BLR test:

- Select pairs $\{x_i, y_i\}_{i=1}^t$ each from $\{0,1\}^n$ independently and u.a.r.
- Accept iff for every $i \in [t]$, $f(x_i) + f(y_i) = f(x_i + y_i)$, otherwise reject.

This test makes $3t$ queries. Clearly if $f$ is a linear function, then the test accepts with probability 1. Show that if the test accepts with probability at least $\frac{1}{2} + \delta$ then $f$ is $\frac{1}{2} + \Omega(\delta)$ correlated with some linear function.

(b) Use the above $t$-fold BLR test (instead of BLR test in the proof of NP-hardness of gap-3LIN($\frac{1}{2} - \epsilon, 1 - \epsilon$)) to construct a PCP verifier for NP with $q = 3t$ queries that accepts a correct proof with probability at least $1 - \epsilon$ and every ‘proof’ of a wrong statement is accepted with probability at most $2^{-t} + \epsilon$, for all $\epsilon > 0$. What is the amortized query complexity of this PCP?
(You will need to modify the $t$-fold BLR test to exclude some linear functions, similar to the modification we did in the proof of NP-hardness of gap-3LIN($\frac{1}{2} - \epsilon, 1 - \epsilon$))

(c) How can we get a PCP with an improved amortized query complexity? In the $t$-fold BLR test, we first query $2t$ locations $x_1, y_1, \ldots, x_t, y_t$. Apart from the $t$ checks, one can try to check if $f(y_i) + f(y_j) = f(y_i + y_j)$, $f(x_i) + f(x_j) = f(x_i + x_j)$ or even $f(x_i) + f(y_j) = f(x_i + y_j)$, for $i \neq j$. Each of these checks needs to query $f$ at only one additional location! Can we hope to reduce the soundness by $1/2$ for each of these checks? So we modify the test as follows:

**Complete Graph Linearity Test:**

- Select $\{x_1, x_2, \ldots, x_t\}$ each from $\{0, 1\}^n$ independently and u.a.r.
- Accept iff for every $i \neq j$, $f(x_i) + f(x_j) = f(x_i + x_j)$, otherwise reject.

Here, we are doing $(\frac{1}{2})^t$ correlated BLR linearity tests. Surprisingly, the soundness of the above test is $2^{-\frac{1}{2}t}$, as if we are performing $(\frac{1}{2})^t$ BLR tests independently! This is what we will prove next.

i. Let $g : \{0, 1\}^n \rightarrow \{-1, +1\}$ be such that $g(x) = (-1)^{f(x)}$. Show that the acceptance probability is

$$\Pr[\text{Accept}] = \frac{1}{2^{\frac{1}{2}t}} + \frac{1}{2^{\frac{1}{2}t}} \cdot \sum_{\emptyset \neq S \subseteq \binom{[t]}{2}} \mathbb{E}_{x_1, x_2, \ldots, x_t} \left[ \prod_{(i,j) \in S} g(x_i)g(x_j)g(x_i + x_j) \right].$$

ii. For any $\emptyset \neq S \subseteq \binom{[t]}{2}$, we want to conclude the following:

$$\mathbb{E}_{x_1, x_2, \ldots, x_t} \left[ \prod_{(i,j) \in S} g(x_i)g(x_j)g(x_i + x_j) \right] \geq \delta \implies \exists T \subseteq [n], \text{ s.t. } |\hat{g}(T)| \geq \delta. \quad (*)$$

Without loss of generality, assume $(1, 2) \in S$. Thus the expression inside the expectation has $g(x_1), g(x_2)$ and $g(x_1 + x_2)$. We will keep these two variables as is and fix the remaining random variables. Show that there exist fixings of $x_3 = a_3, x_4 = a_4, \ldots, x_t = a_t$ such that

$$\mathbb{E}_{x_1, x_2, \ldots, x_t} \left[ \prod_{(i,j) \in S} g(x_i)g(x_j)g(x_i + x_j) \right].$$
\[
\leq \mathbb{E}_{x_1, x_2} \left[ g(x_1)g(x_2)g(x_1 + x_2) \prod_{(1, j) \in S, \ j \neq 2} g(x_1)g(a_j)g(x_1 + a_j) \prod_{(i, 2) \in S, \ i \neq 1} g(a_i)g(x_2)g(a_i + x_2) \right]
\]

iii. If we define functions, \( h : \{0, 1\}^n \to \{-1, +1\} \) and \( h' : \{0, 1\}^n \to \{-1, +1\} \) as
\[
h(z) \overset{\text{def}}{=} g(z) \prod_{(1, j) \in S, \ j \neq 2} g(z)g(a_j)g(z + a_j)
\]
and
\[
h'(z) \overset{\text{def}}{=} g(z) \prod_{(i, 2) \in S, \ i \neq 1} g(a_i)g(z)g(a_i + z),
\]
then from the assumption (⋆) and (ii), conclude
\[
\left| \mathbb{E}_{x_1, x_2} [h(x_1)h'(x_2)g(x_1 + x_2)] \right| \geq \delta. \quad (⋆⋆)
\]

iv. From (⋆⋆), using analysis similar to the analysis of the BLR test, conclude that there exists \( T \subseteq [n], \) s.t. \( |g(T)| \geq \delta. \)

(d) Use the Complete Graph Linearity Test to construct a PCP with \( q \)-queries, completeness \( 1 - \epsilon \), and amortized query complexity \( 1 + o_q(1) \).