Homework 11

Problem 1. Prove that the function family

$$\mathcal{H} = \left\{h_{a,b} \mid h_{a,b}(x) = a \cdot x + b, a \in \{0,1\}^k, b \in \{0,1\}\right\}$$

is a pairwise independent hash function family for range $R = \{0,1\}$ and domain $U = \{0,1\}^k$.

Solution. All the computations are under modulo 2.

> **Lemma.** Given $x \in \{0,1\}^k, x \neq 0$, then
>
> $$\Pr_{a \in \{0,1\}^k}(a \cdot x = 0) = \frac{1}{2}.$$
>
> **Proof.** Assume $x_i = 1$ for some $i \in [0, k-1]$. Then
>
> $$\Pr_{a \in \{0,1\}^k}(a \cdot x = 0) = \Pr_{a_i \in \{0,1\}}(a_i = \sum_{j \neq i} a_j x_j) = \frac{1}{2}.$$
>
> □

It is easy to verify that

$$\Pr_{a \in \{0,1\}^k, b \in \{0,1\}}(h(x_1) = y_1) = \frac{1}{|R|} = \frac{1}{2},$$

so we only need to prove that, for any two distinct elements $x_1, x_2 \in U = \{0,1\}^k$, and two (possibly equal) elements $y_1, y_2 \in R = \{0,1\}$, we have

$$\Pr_{a \in \{0,1\}^k, b \in \{0,1\}}(h_{a,b}(x_1) + b = y_1 \text{ and } h_{a,b}(x_2) + b = y_2) = \frac{1}{4}.$$

This is equivalent to proving that

$$\Pr_{a \in \{0,1\}^k, b \in \{0,1\}}(a \cdot x_1 + b = y_1 \text{ and } a \cdot x_2 + b = y_2) = \frac{1}{4}.$$

Separate a and b as follows:

$$\begin{aligned}
&\Pr_{a \in \{0,1\}^k, b \in \{0,1\}}(h_{a,b}(x_1) = y_1 \text{ and } h_{a,b}(x_2) = y_2) \\
&= \Pr_{a,b}(a \cdot x_1 + b = y_1 \text{ and } a \cdot x_2 + b = y_2) \\
&= \Pr_{a,b}(a \cdot (x_1 \oplus x_2) = y_1 \oplus y_2 \text{ and } b = y_1 \oplus a \cdot x_1) \\
&= \Pr_{a}(a \cdot (x_1 \oplus x_2) = y_1 \oplus y_2) \cdot \Pr_{b}(b = y_1 \oplus a \cdot x_1 \mid a = a_0).
\end{aligned}$$

Now that $x_1 \oplus x_2 \neq 0$, so by the lemma, we know that

$$\Pr_a(a \cdot (x_1 \oplus x_2) = y_1 \oplus y_2) = \frac{1}{2}.$$

Also, fix $a = a_0$, then $\Pr_b(b = y_1 \oplus a \cdot x_1 \mid a = a_0) = \frac{1}{2}$. Therefore we have

$$\Pr_{a \in \{0,1\}^k, b \in \{0,1\}}(a \cdot x_1 + b = y_1 \text{ and } a \cdot x_2 + b = y_2) = \frac{1}{4},$$

and $\mathcal{H}$ is a pairwise independent hash function family.

Problem 2.

(a) Consider a random walk $X_0, X_1, X_2, \ldots$ on a chain of $n+1$ vertices $0, 1, \ldots, n$ with the following transition probabilities

$$\Pr(X_t = k | X_{t-1} = j) = \begin{cases} \frac{1}{2} & \text{if } j \in [1, n-1] \text{ and } k = j \pm 1, \\ 1 & \text{if } j = 0, k = 1 \text{ or } j = n, k = n, \\ 0 & \text{otherwise.} \end{cases}$$

Let T_i be the expected number of steps the walk takes to arrive at the end vertex n starting with $X_0 = i$. Prove that $T_i \leq n^2$ for all $i \in [0, n]$.

(b) Consider the following randomized algorithm for 2-SAT problems of n variables.

```
Choose an arbitrary initial assignment.
for t = 1, 2, ..., 2n^2 do
    if the current assignment is satisfying then
        Accept immediately.
    else
        Choose an arbitrary clause not satisfied.
        Sample one of the two literals uniformly at random.
        Flip the value of the variable in the sampled literal.
    end if
end for
Reject if haven't accepted.
```

Use Markov inequality to show that the algorithm will find a satisfying solution with probability at least $\frac{1}{2}$ given a yes-instance as input.

Solution. (a)

From the definition of T_i, we have

$$T_i = 1 + \frac{1}{2}T_{i-1} + \frac{1}{2}T_{i+1}, \quad 1 \le i \le n-1,$$

and

$$T_0 = 1 + T_1, \quad T_n = 1.$$

We can rewrite the above equations as

$$T_i - T_{i-1} = T_{i+1} - T_i + 2 \quad 1 \le i \le n-1.$$

Summing up the above equations for $i \in [1, n-1]$, we have

$$T_{n-1} - T_0 = T_n - T_1 + 2(n-1).$$

Since $T_0 = 1 + T_1$ and $T_n = 1$, we have

$$T_{n-1} = 2n - 1.$$

Now we will prove that $T_i = n^2 - i^2$ holds for all $i \in [0, n]$ by induction.

- **Base case:** $T_n = n^2$ and $T_{n-1} = n^2 - (n-1)^2 = 2n - 1$.
- **Inductive step:** Suppose that $T_i = n^2 - i^2$ holds for all $i \ge k \in [1, n-1]$. Then we have $T_{i-1} = 2T_i - T_{i+1} - 2 = 2(n^2 - i^2) - (n^2 - (i+1)^2) - 2 = n^2 - (i-1)^2$.
- **Conclusion:** By induction, we have $T_i = n^2 - i^2$ for all $i \in [1, n]$. Consider the special case $i = 0$, we have $T_0 = 1 + T_1 = 1 + (n^2 - 1) = n^2$. Therefore, we have $T_i = n^2 - i^2$ for all $i \in [0, n]$.

Hence, $T_i \le n^2$ for all $i \in [0, n]$.

(b)

Consider the random variable X_i. X_i is the number of steps used to reach the state that i variables have the correct assignments. Suppose we have sampled the literal $l_{i,1}$.

- If it already has the correct value, then after we flip its value, it would be wrong, ending up with the state X_{i-1}.
- If it has the wrong value, then after we flip its value, it would be correct, ending up with the state X_{i+1}.

Each of the cases above has a probability of $\frac{1}{2}$. Therefore, it is the same procedure as the random walk in part (a).

The target is to have all the n variables assigned with a correct value. And the number of steps to reach the state that all the variables have the correct assignments (i.e. the state X_n), from the state that initially have i correct assignments (i.e. the state X_i), is Y_i. Denote $T_i = \mathbb{E}(Y_i)$ as its expectation.

Then, by Markov inequality, we have

$$\Pr(Y_i \geq 2n^2) \leq \frac{\mathbb{E}(Y_i)}{2n^2} = \frac{T_i}{2n^2} \leq \frac{n^2}{2n^2} = \frac{1}{2}.$$

Hence, after $2n^2$ steps, the algorithm will find a satisfying solution with probability at least $\frac{1}{2}$ given a yes-instance as input.