6. Error reduction with pairwise independent hash family

6.1 Introduction

In the first part of this lecture, we will see how to do randomness-efficient error reduction using pairwise independent hash families. Following that we will start with the first "pseudorandom" object that we will meet in this course - expanders.

6.2 Pairwise independent hash family

Definition 6.1 (Pairwise independent hash family)**.** *A set $\mathcal{H} = \{h : [N] \to [M]\}$ of functions is a pairwise independent hash family if it satisfies the following two conditions: For every $x_1 \neq x_2 \in [N]$ and $y_1, y_2 \in [M]$, $\Pr_{h \in_r \mathcal{H}}[H(x_1) = y_1 \wedge H(x_2) = y_2] = \frac{1}{M^2}$.*

Definition 6.2. *Let $m < n$. For $a \in \{0,1\}^n, b \in \{0,1\}^m$, let $h_{a,b} : \{0,1\}^n \to \{0,1\}^m$ be defined as $h_{a,b}(x) = (ax)|_m + b$, where $(y)|_m$ is the string y truncated to m bits. Let $\mathcal{H} = \{h_{a,b} \mid a, b \in \mathbb{F}\}$.*

Here, we think of a and x as elements of the field $\mathbb{F}_{2^n}$, perform the multiplication in the field $\mathbb{F}_{2^n}$ polynomials and map them back to binary strings. This hash family can be represented by $m + n$ bits.

6.2.1 Application to error reduction

Lemma 6.3. *Let $X_1, X_2, \ldots, X_t$ be pairwise independent random variables taking values in the interval $[0,1]$, and let $X = \sum_i X_i$ and $\mu = E[X]$. Then,*

$$\Pr[|X - \mu| \geq k] \leq \frac{t}{k^2}$$

Proof.

$$\begin{aligned}
\mathrm{Var}[X] = E[(X-\mu)^2] &= E\left[\left(\sum_i (X_i - \mu_i)\right)^2\right] \\
&= E\left[\sum_{i,j} (X_i - \mu_i)(X_j - \mu_j)\right] \\
&= \sum_{i,j} E[(X_i - \mu_i)(X_j - \mu_j)] \\
&= \sum_i E[(X_i - \mu_j)^2] \leq t
\end{aligned}$$

Now apply Chebyshev's inequality. □

For a randomized algorithm that has an error probability of $1/2 - 1/n$, to improve the error probability to e^{-6}, we needed about $10n^2$ runs of the algorithm each with an independent set of random bits. When using the bound above, we need around $e^6 n^2$ rounds to get an error bound of e^6. But now, the random bits that are used in each iteration need to be only pairwise independent. So, if the original algorithm need $r(n)$ bits of randomness, for the new algorithm, I need just $r(n) + \log n$ bits of randomness.

Exercise 6.4. *Let X be a random variable taking values in $[0, 1]$ and let $\mu = E[X]$. Then show that $\mathrm{Var}[X] \leq \mu(1 - \mu)$.*

6.3 Expander graphs

Now we will look at expander graphs, sparse graphs that are highly connected. There are multiple ways to define what expansion is in a graph. We start with the definition of vertex expansion. We will mostly be looking at directed multigraphs (undirected graphs can be thought of as directed graphs with edges in both direction). We will mostly be looking at regular graphs.

Definition 6.5. *A digraph is a (K, A) vertex expander if for all subsets S of the vertices such that $|S| \leq K$, the neighborhood of S defined as $N(S) = \{u \mid \exists v \ s.t \ (u, v) \in E\}$ is of size at least $A \cdot |S|$.*

This definition of expansion is not unique, and it can be defined in multiple ways. We will see these definitions along the way and also see that they are almost equivalent.

Lemma 6.6. *For every constant $D \geq 32$, there exists a constant $n_0 > 0$ such that for all $N > n_0$, there exists a $(\frac{n}{10D}, \frac{5D}{8})$ vertex expander.*

Example 6.7. *For every integer m, G_m is a graph on the vertex set $\mathbb{Z}_m \times \mathbb{Z}_m$, and the edge set is given as follows: (x, y) is connected to $(x + y, y), (x - y, y), (x, y + x), (x, y - x), (x + y + 1, y), (x - y + 1, y), (x, y + x + 1), (x, y - x + 1)$. This construction due to Margulis is the first known explicit expander family.*

In the next lecture we will see how to amplify the success probability of one-sided error algorithms. We will then look at the notion of spectral expansion and its relation to vertex expansion.