45. Small-bias spaces

45.1 Introduction

Now we study small-bias spaces, their explicit constructions, properties and some applications. This note contains materials that were covered in Lectures 43,44, and 45.

45.2 ε-bias spaces

Definition 45.1 (ε-bias distribution)**.** *A distribution μ over $\{0,1\}^n$ is an ε-bias distribution if for every set $S \subseteq [n]$, we have the following:*

$$\left| \Pr_{x\sim\mu}\left[\sum_{i\in S} x_i = 1\right] - \Pr_{x\sim\mu}\left[\sum_{i\in S} x_i = 0\right]\right| \leq \varepsilon.$$

A set $S \subseteq \{0,1\}^n$ is an ε-bias space if the uniform distribution over S is an ε-bias distribution.

If we think of the distribution μ as a function $\mu : \{0,1\}^n \to [0,1]$, we can show that $|\hat{\mu}(S)| \leq 2^{-n}\varepsilon$ for every non-empty $S \subseteq [n]$. Observe that for the uniform distribution $\mathcal{U}_n$, $\hat{\mathcal{U}_n}(S) = 0$ for all $S \neq \emptyset$ and $\hat{\mathcal{U}_n}(\emptyset) = \frac{1}{2^n}$. Therefore we have the following simple observation.

Lemma 45.2. *If μ is an ε-bias distribution over $\{0,1\}^n$, then $\|\mu - \mathcal{U}_n\|_2 \leq \varepsilon$.*

Proof. First observe that $\hat{\mu}(\emptyset) = \frac{1}{2^n}\sum_x \mu(x) = \frac{1}{2^n}$. Then we have

$$\begin{aligned}\|\mu - \mathcal{U}_n\|_2^2 &= 2^n \sum_{S\subseteq[n]} \left(\hat{\mu}(S) - \hat{\mathcal{U}_n}(S)\right)^2 \\ &= 2^n \sum_{S\neq\emptyset} \hat{\mu}(S)^2 \leq 2^n 2^{-2n}\varepsilon^2(2^n - 1) \leq \varepsilon^2.\end{aligned}$$

□

We now define the notion of pseudorandom distributions.

Definition 45.3 (PRG)**.** *A distribution μ is ε-pseudorandom for a class $\mathcal{C}$ of Boolean functions, if for every $f \in \mathcal{C}$, we have*

$$\left| \Pr_{x\sim\mu}[f(x) = 1] - \Pr_{x\sim\mathcal{U}_n}[f(x) = 0]\right| \leq \varepsilon.$$

The following observation about small-bias spaces is easy to see.

Lemma 45.4. *An ε-bias distribution μ is $\varepsilon/2$-pseudorandom for the class of linear functions on $\mathbb{F}_2^n$.*

Proof. Since $\Pr_{x\sim\mu}\left[\sum_{i\in S} x_i = 1\right] + \Pr_{x\sim\mathcal{U}_n}\left[\sum_{i\in S} x_i\right] = 1$, for every $S \subseteq [n]$, we have

$$\left|\Pr_{x\sim\mu}\left[\sum_{i\in S} x_i = 1\right] - \frac{1}{2}\right| = \left|\Pr_{x\sim\mu}\left[\sum_{i\in S} x_i = 1\right] - \Pr_{x\sim\mathcal{U}_n}\left[\sum_{i\in S} x_i = 1\right]\right| \leq \frac{\varepsilon}{2}$$

□

45.3 Balanced codes and small-bias spaces

The notion of small-bias spaces are closely connected to balanced linear codes.

Definition 45.5. *A binary linear code $\mathcal{C} = [n, k, d]_2$ is an* ε-weight balanced *code if every non-zero codeword $x \in \mathcal{C}$ has Hamming weight* $\mathrm{wt}(x) \in \left[\frac{1-\varepsilon}{2}n, \frac{1+\varepsilon}{2}n\right]$.

The Walsh-Hadamard code is an ε-weight balanced code since every non-zero codeword has weight exactly $n/2$. We will now show that balanced codes give small-bias spaces.

Lemma 45.6. *Let $\mathcal{C} = [n, k, d]_2$ be an ε-weight balanced binary code, and let $G \in \mathbb{F}_2^{k\times n}$ be its generator matrix. Then, the set of columns of G is an ε-bias space.*

Proof. Let $S = C_1, C_2, \ldots, C_n$ be the set of columns of G and each column $C_i \in \mathbb{F}_2^k$. Let μ be the uniform distribution over S. To prove the lemma, we need to show that for every $T \subseteq [k]$ we have

$$\left|\Pr_{C_j\sim\mu}\left[\sum_{i\in T} C_{i,j} = 1\right] - \Pr_{C_j\sim\mu}\left[\sum_{i\in T} C_{i,j} = 0\right]\right| \leq \varepsilon.$$

To that end, fix $T \subseteq [k]$. This defines a vector $\chi_T \in \mathbb{F}_2^k$ which is the characteristic vector of T. Now $\chi_T \cdot G$ gives a vector $y \in \mathcal{C}$ such that $\mathrm{wt}(y) \in \left[\frac{1-\varepsilon}{2}n, \frac{1+\varepsilon}{2}n\right]$. Then, $\Pr_{i\in_r[n]}[y_i = 1] \in \left[\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}\right]$. So, we have $|\Pr_{i\in_r[n]}[y_i = 1] - \Pr_{i\in_r[n]}[y_i = 0]| \leq \varepsilon$. But, $\Pr_{i\in_r[n]}[y_i = 1] = \Pr_{C_j\sim\mu}\left[\sum_{i\in T} C_{i,j} = 1\right]$. □

45.3.1 Small-bias spaces from concatenated codes

We will briefly explain how to get explicit small-bias spaces from concatenated codes. Consider a Reed-Solomon code $\mathcal{C}_1$ with parameters $[q, k_1, q - k_1]_q$ where $q = \frac{k_1}{\varepsilon}$. Let $\mathcal{C}_2$ be a Walsh-Hadamard code with parameters $[q, \log q, \frac{q}{2}]_2$. The concatenation of $\mathcal{C}_1$ and $\mathcal{C}_2$ is obtained as follows. Let $x \in [q]^{k_1}$ be the message. First we obtain a Reed-Solomon codeword y of length q, where each $y_i \in [q]$. We encode each y_i with a $\log q$-bit binary string. Now we encode each y_i using the Walsh-Hadamard code $\mathcal{C}_2$ to obtain $z_i \in \mathbb{F}_2^q$. The final codeword z is the concatenation of z_is and has length q^2.

The concatenated code takes binary strings of length $k_1 \log q$ and obtaines codewords of length $q^2 = \left(\frac{k_1}{\varepsilon}\right)^2$. Every non-zero string y obtained from the Reed-Solomon encoding has at least $1-\varepsilon$ fraction of non-zero entries. Each non-zero entry is converted by the Walsh-Hadamard code into a string with $1/2$ fraction of ones. So the concatenated code has relative weight at least $\frac{1-\varepsilon}{2}$. Since every non-zero Walsh-Hadamard codeword has relative weight exactly $1/2$, the relative weight of the concatenated codeword is at most $1/2$ and hence this is an ε-balanced code. This gives an ε-bias space over $\mathbb{F}_2^k$ (where $k = k_1 \log q$) of size at most k^2/ε^2.

45.4 Small-bias spaces and Cayley expanders

Let G be a group and $S \subseteq G$ any subset of the group. The Cayley graph $\mathrm{Cay}(G, S)$ is a grph whose vertex set is the set of elements of the group G. A vertex $g \in G$, has an edge $(g, g \cdot s)$ for every $s \in S$ and $\cdot$ is the group operation of G. We will now see that for $G = \mathbb{F}_2^n$, an ε-bias space S gives a Cayley graph $\mathrm{Cay}(G, S)$ that is an ε-spectral expander.

Let A be the normalized adjacency matrix of the Cayley graph $\mathrm{Cay}(G, S)$. The matrix A is a $2^n \times 2^n$ matrix where the rows and columns are indexed by the elements of the group G. For a row indexed by an element $g \in \mathbb{F}_2^n$, for each $s \in S$, the entry $A[g, g \cdot s] = \frac{1}{|S|}$. Let χ_T denote the linear function $\chi_T(x) = \sum_{i \in T} x_i$. We can think of χ_T as 2^n-bit vector. We will show that each χ_T is an eigenvector of A with eigenvalue at most ε.

For an element $g \in \mathbb{F}_2^n$, we have

$$A \cdot \chi_T(g) = \sum_{s \in S} \frac{1}{|S|} \chi_T(g \cdot s) = \sum_{s \in S} \frac{1}{|S|} \chi_T(g) \chi_T(s) = \chi_T(g) \sum_{s \in S} \frac{1}{|S|} \chi_T(s).$$

Therefore χ_T is an eigenvector with eigencalue $\sum_{s \in S} \frac{1}{|S|} \chi_T(s)$. Let μ be the uniform distribution over S. Then we have

$$\left| \sum_{s \in S} \frac{1}{|S|} \chi_T(s) \right| = \left| \mathbb{E}_{s \sim \mu} \left[\sum_{i \in T} s_i \right] \right| = \left| \Pr_{s \sim \mu} \left[\sum_{i \in T} s_i = 1 \right] - \Pr_{s \sim \mu} \left[\sum_{i \in T} s_i = 0 \right] \right| \leq \varepsilon.$$