CS761 Derandomization and Pseudorandomness **2022-23 Sem I**

Lecture 19: 17-10-2022

Scribe: Amit Rajaraman *Lecturer: Rohit Gurjar*

In the last lecture, we saw that the relative distance of the Reed Muller code was $1 - d/|\mathbb{F}|$, when viewed as a code on alphabet $\mathbb{F}$. When viewed as a code on alphabet $\{0,1\}$ however, this goes to $(1 - d/|\mathbb{F}|)/\log|\mathbb{F}|$. This issue of the relative distance being $o(1)$ cannot be fixed even by changing $\mathbb{F}, \ell, d$.
To fix this, we shall do the following: for each element of $\mathbb{F}$ (each coordinate when viewed as a code on alphabet $\mathbb{F}$), we shall replace it with another codeword, possibly larger. That is, if we encode it as $x \in \mathbb{F}^{|\mathbb{F}|^\ell}$ under the Reed-Muller code, we encode each x_i as another element $\{0,1\}^t$, where t will end up being $|\mathbb{F}|$.

This second code is the *Walsh-Hadamard code*, defined as follows. The encoding is a function WH : $\{0,1\}^k \to \{0,1\}^{2^k}$, where for each $S \subseteq [k]$, we have $(\mathrm{WH}(x))_S = \bigoplus_{i\in S} x_i$.
We claim that the relative distance of this code is $1/2$. Indeed, if we change r bits, all coordinates corresponding to subsets that contain an odd number of these r bits will change.
Further, it turns out that the Walsh-Hadamard code is optimal.

Proposition 19.1. *Let $E : \{0,1\}^n \to \{0,1\}^m$ be a code with $m < 2^n - 1$. Then, the relative distance of E is at most $1/2$.*

Proof sketch. Suppose instead that E is a function to $\{-1,1\}^m$ (replacing 0 with -1) with relative distance $\Delta > (1/2)$. Note that $\langle f(x), f(y)\rangle < 0$ for any distinct $x, y \in \{0,1\}^n$. The number of such vectors is at most $m + 1 < 2^n$ (see, for example, here) so we are done. □

In addition, the Walsh-Hadamard code is locally decodable. Given x and some corruption of $\mathrm{WH}(x)$, we can consider sets of the form T and $T \cup \{i\}$, where $i \notin T$. Adding (XORing) the two should give x_i in the absence of corruption. When there is corruption, we can just choose a bunch of random T and perform this same operation, taking the majority finally. The probability that either both $\mathrm{WH}(x)_T$ and $\mathrm{WH}(x)_{T\cup\{i\}}$ are uncorrupted or both are corrupted (since we get the correct value of x_i in either case) is $(1-\rho)^2 + \rho^2 = 1 - 2\rho(1-\rho)$. When $\rho < (1/2)$, this is more than $1/2$ so the majority value gives the correct value with high probability.

In conclusion, our final code is $\mathrm{WH}(\mathrm{RM}(x))$.[1] Here, WH is a mapping from $\{0,1\}^{\log|\mathbb{F}|} \to \{0,1\}^{|\mathbb{F}|}$. The relative distance of this code is $(1/2)(1 - d/|\mathbb{F}|)$, which is $\Theta(1)$ for appropriate $d, |\mathbb{F}|$! We can handle an error fraction of $\rho \approx \Delta/2 \approx (1/4)$.
One interesting thing is that due to the previous proposition, we cannot even do better than $1/4$ using a coding theory-based proof like this.

Now, we have gone from exponential H_{worst} to exponential $H_{\text{avg}}^{1-\rho}$, which in the limiting case is $H_{\text{avg}}^{3/4}$. How do we go from this to H_{avg}? We do not delve into the details of this, but the main result used is the following.

Theorem 19.2 (Yao's XOR Lemma)**.** *Given a function $f : \{0,1\}^n \to \{0,1\}$, define the function $\hat{f} : \{0,1\}^{nk} \to \{0,1\}$ defined by*

$$f(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k) = f(\overline{x}_1) \oplus f(\overline{x}_2) \oplus \cdots \oplus f(\overline{x}_k),$$

where each $\overline{x}_i$ is in $\{0,1\}^n$.
If $\delta > 0$ and $\epsilon > 2(1-\delta)^k$,

$$H_{avg}^{(1/2)+\epsilon}(\hat{f}) \geq \frac{\epsilon^2}{400n} H_{avg}^{1-\delta}(f).$$

[1] mildly abusing notation to mean that we apply WH on a coordinate-by-coordinate basis to $\mathrm{RM}(x)$.

Given a function with exponentially large $H_{\text{avg}}^{1-\delta}$, making ϵ appropriately exponentially small (about $H_{\text{avg}}^{1-\delta}(f)^{-1/3}$) does the job.
Alternatively, one way to go directly from H_{worst} to H_{avg} is to use *local list decoding* for the Reed-Muller and Walsh-Hadamard combination we saw earlier in the lecture.