6.5420 Randomness and Computation April 16, 2026

Lecture 17

Lecturer: Ronitt Rubinfeld *Scribe: Brandon Lou*

1 Topics

- Fourier concentration via noise sensitivity
- Learning heavy Fourier coefficients (with queries)

2 Fourier Background

Recall:

- $\hat{f}(S) = \mathbb{E}[f(x)\chi_S(x)]$
- $\hat{f}(S) = \langle f, \chi_S \rangle = 1 - 2 \cdot Pr[f(x) \neq \chi_S(x)]$
- Parseval's identity for Boolean:

$$\sum_S \hat{f}(S)^2 = \langle f, f \rangle = 1$$

- We can estimate any Fourier coefficient $\hat{f}(S)$ up to additive error γ with probability $1-\delta$ using $O(\frac{1}{\gamma^2}\log(1/\delta))$ samples.

3 Noise Sensitivity

Definition 1 (Noise Operator)**.** *Let $N_\varepsilon(x)$ be obtained by flipping each bit of x independently with probability ε.*

Definition 2 (Noise Sensitivity)**.**

$$NS_\varepsilon(f) = \Pr_{x \in \{\pm 1\}^n + noise}[f(x) \neq f(NS_\varepsilon(x)]$$

We state the following theorem without proof (proof will be done for homework):

Theorem 3. *For $f : \{\pm 1\}^n \to \{\pm 1\}$,*

$$NS_\varepsilon(f) = \frac{1}{2} - \frac{1}{2}\sum_S (1-2\varepsilon)^{|S|}\hat{f}(S)^2$$

3.1 Examples

Below are examples of noise sensitivities for functions that we have seen previously:

- For the dictator function $f(x) = x_1$:

$$NS_\varepsilon(f) = \varepsilon$$

- For the parity function χ_T:

$$NS_\varepsilon(\chi_T) = \frac{1}{2} - \frac{1}{2}(1-2\varepsilon)^{|T|}\hat{f}(T)^2 NS_\varepsilon(\chi_T) = \frac{1}{2} - \frac{1}{2}(1-2\varepsilon)^{|T|}$$

  since $\hat{f}(T)^2 = 1$ for parity functions.

- Linear threshold functions (LTFs):

$$NS_\varepsilon(\text{LTF}) \leq 8.8\sqrt{\varepsilon}$$

4 Noise Sensitivity ⇒ Fourier Concentration

We can use theorem 3 to relate noise sensitivity to Fourier concentration.

Theorem 4.

$$\sum_{|S|\geq 1/\gamma} \hat{f}(S)^2 \leq \frac{2}{1-e^{-2}} \cdot NS_\gamma(f) \leq 2.32 \cdot NS_\gamma(f)$$

This tells us how much we can ignore Fourier coefficients that are bigger than $1/\gamma$.

Proof.

$$2NS_\gamma(f) = 1 - \sum_S (1-2\gamma)^{|S|}\hat{f}(S)^2$$

Using Parseval:

$$= \sum_S \hat{f}(S)^2 \left[1 - (1-2\gamma)^{|S|}\right]$$

Restricting to $|S| \geq 1/\gamma$:

$$\geq \sum_{|S|\geq 1/\gamma} \hat{f}(S)^2 \left[1 - (1-2\gamma)^{1/\gamma}\right]$$

Using $(1-2\gamma)^{1/\gamma} \leq e^{-2}$:

$$\geq (1-e^{-2}) \sum_{|S|\geq 1/\gamma} \hat{f}(S)^2$$

Rearranging gives the result. □

What this theorem tells us is that if we have a good bound on noise sensitivity, then we have a bound on Fourier concentration.

Corollary 5. *For an LTF,*

$$\sum_{|S|\geq O(1/\varepsilon^2)} \hat{f}(S)^2 \leq \varepsilon$$

This is an immediate corollary of the theorem, where we take $\gamma = \varepsilon^2$. We also rewrite the equation to get rid of the constant coefficient on the right hand side.

Implication:

- LTFs can be learned from $n^{O(1/\varepsilon^2)}$ samples

- Functions of k LTFs can be learned with $n^{O(k^2/\varepsilon^2)}$ samples.

  Suppose that $f(x) = g(f_1(x), f_2(x), \dots f_k(x))$, where each f_i is an LTF. We start with the bound for noise sensitivity of a single halfspace, $NS_\varepsilon(f_i) = O(\sqrt{\varepsilon})$. We can now union bound the probability that we flip any of the halfspaces. We have

  $$\Pr[\text{any } f_i \text{ flips}] \leq \sum_{i=1}^{k} NS_\varepsilon(f_i) = k \cdot O(\sqrt{\varepsilon})$$

  This bounds the noise sensitivity of f to be k times the noise sensitivity of one halfspace:

  $$NS_\varepsilon(f) = O(k\sqrt{\varepsilon})$$

Halfspaces are an example of low noise sensitivity. In the next section, we will see what happens when we no longer have low noise sensitivity.

5 Learning Heavy Fourier Coefficients

Parity functions have high noise sensitivity:

$$NS_\varepsilon(f) \approx \frac{1}{2}$$

Thus, they do *not* exhibit Fourier concentration and cannot be learned via low-degree methods.

However, parity functions are still learnable in a different model. Suppose we are given query access to a parity function $f(x) = \chi_S(x)$. We can collect random labeled examples $(x^{(i)}, y_i)$ and solve a linear system:

$$\begin{pmatrix} x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \\ \vdots & & \vdots \\ x_{m1} & \dots & x_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

If there is no noise and we have enough samples, we can recover the unknown coefficients a_i using linear algebra.

However, this approach is fragile:

- It relies on the function being exactly linear
- There has to be no noise. Otherwise, we can no longer rely on linear algebra.

For example, consider the function:

- if $x_1 = +1$, output $+1$ ($\chi_\emptyset$)
- if $x_1 = -1$, output $\prod_i x_i$ ($\chi_{[n]}$)

This function is a combination of linear functions: it combines a low-degree component and a high-degree parity component, but the function itself is non-linear, and so our previous methods of recovering this function do not work. We will introduce a method which will work for learning heavy Fourier coefficients.

Goal (Goldreich-Levin / Kushilevitz-Mansour):

- Output all S such that $|\hat{f}(S)| \geq \theta$ (all "close" linear functions)
- Only output S with $|\hat{f}(S)| \geq \theta/2$ (no junk)

We assume **query access** to f.

5.1 Main Idea

The main idea is to use an exhaustive search, but with good pruning, to avoid having to search every possible set.

- Build a binary tree over variables
- Each level corresponds to a a variable X_i. Branch left if $X_i \in S$, and branch right if $X_i \notin S$.
- Each node corresponds to partial set $S_1 \subseteq [k]$. The leaves contain every possible set of coefficients.
- Recursively extend by including/excluding the next variable
- Prune subtrees with low "energy". We create an algorithm to estimate the total energy of a subtree, and only traverse if that energy is high enough.

5.2 Key Definition

If we fix $0 \leq k \leq n$ (the current level), then $S_1 \subseteq [k]$ tells us what node we reached, as well as which of the previous nodes we traversed. In other words, it tells us for the first k variables, which are in or out.

We also define, for $S_1 \subseteq [k]$, a function $f_{k,S_1} : \{\pm 1\}^{n-k} \to \mathbb{R}$, such that

$$f_{k,S_1}(x) = \sum_{T_2 \subseteq \{k+1,\ldots,n\}} \hat{f}(S_1 \cup T_2)\chi_{T_2}(x)$$

Note that we are summing over T_2, which is how much energy is left in the Fourier coefficients where we begin with S_1.

5.3 Sanity Checks

- $k = 0$. At the root, we have

  $$f_{0,\emptyset} = \sum_{T_2 \subseteq [n]} \hat{f}(T_2)\chi_T(x) = f(x)$$

  since we have $k = 0$ and $S_1 = \emptyset$. So, at the root, we should get back our function.

- $k = n$: At a leaf, we have

  $$f_{n,S_1} = \hat{f}(S_1)\chi_\emptyset(x) = \hat{f}(S_1)$$

  since we are summing over $T_2 = \emptyset$.

5.4 Pruning Criterion

Only recurse if:

$$\mathbb{E}[f_{k,S_1}(x)^2] \geq \theta^2$$

But, we need to figure out, can we estimate this expected value, and can we get to the right leaves (heavy leaves)? Also, how much junk do we have, and how many paths do we take?

For this lecture, we will see that the answer to the last question is that we don't take too many paths.

5.5 Key Lemma ("Not Too Many")

Lemma 6. *The "not too many" lemma states:*

- *At most $\frac{1}{\theta^2}$ sets S satisfy $|\hat{f}(S)| \geq \theta$*
- *At each level k, at most $\frac{1}{\theta^2}$ nodes satisfy:*

$$\mathbb{E}[f_{k,S_1}(x)^2] \geq \theta^2$$

Proof. First part follows from Boolean Parseval's. If more than $1/\theta^2$ S's satisfy $|\hat{f}(S)| \geq \theta$, then $\sum_S \hat{f}(S)^2 > 1$, which is a contradiction.

For the second part, fix k and a subset $S_1 \subseteq [k]$. We first prove the identity:

$$\mathbb{E}_x[f^2_{k,S_1}(x)] = \sum_{T_2 \subseteq \{k+1,\ldots,n\}} \hat{f}(S_1 \cup T_2)^2.$$

By definition,

$$f_{k,S_1}(x) = \sum_{T_2 \subseteq \{k+1,\ldots,n\}} \hat{f}(S_1 \cup T_2)\chi_{T_2}(x).$$

Therefore,

$$\mathbb{E}_x[f^2_{k,S_1}(x)] = \mathbb{E}_x\left[\left(\sum_{T_2} \hat{f}(S_1 \cup T_2)\chi_{T_2}(x)\right)^2\right].$$

Expanding the square,

$$= \mathbb{E}_x\left[\sum_{T_2,T_2'} \hat{f}(S_1 \cup T_2)\hat{f}(S_1 \cup T_2')\chi_{T_2}(x)\chi_{T_2'}(x)\right].$$

By linearity of expectation,

$$= \sum_{T_2,T_2'} \hat{f}(S_1 \cup T_2)\hat{f}(S_1 \cup T_2')\,\mathbb{E}_x[\chi_{T_2}(x)\chi_{T_2'}(x)].$$

Using orthonormality of the Fourier characters,

$$\mathbb{E}_x[\chi_{T_2}(x)\chi_{T_2'}(x)] = \begin{cases} 1 & \text{if } T_2 = T_2', \\ 0 & \text{otherwise,} \end{cases}$$

so all cross terms vanish and we obtain

$$\mathbb{E}_x[f^2_{k,S_1}(x)] = \sum_{T_2} \hat{f}(S_1 \cup T_2)^2.$$

Summing over all $S_1 \subseteq [k]$, we get

$$\sum_{S_1} \mathbb{E}_x[f^2_{k,S_1}(x)] = \sum_{S_1}\sum_{T_2} \hat{f}(S_1 \cup T_2)^2 = \sum_S \hat{f}(S)^2.$$

By Parseval's identity,

$$\sum_S \hat{f}(S)^2 = 1.$$

Thus,

$$\sum_{S_1} \mathbb{E}_x[f^2_{k,S_1}(x)] = 1.$$

It follows that there can be at most $1/\theta^2$ subsets S_1 such that

$$\mathbb{E}_x[f^2_{k,S_1}(x)] \geq \theta^2,$$

since otherwise the sum would exceed 1, contradicting Parseval's identity. □