6.5420 Randomness and Computation April 14, 2026

Lecture 18

Lecturer: Ronitt Rubinfeld *Scribe: Fedir Yudin*

1 Content

- Low-degree algorithm wrap-up
- Fourier Concentration
- Noise Sensitivity

2 Review

Recall the Fourier basis on $\{\pm 1\}^n$:

$$\chi_S(x) = \prod_{i \in S} x_i.$$

The inner product is

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x)g(x).$$

The Fourier coefficient of f on S is

$$\hat{f}(S) = \langle f, \chi_S \rangle.$$

For Boolean-valued $f : \{\pm 1\}^n \to \{\pm 1\}$, one can also write

$$\hat{f}(S) = 1 - 2\Pr[f(x) \neq \chi_S(x)].$$

Every function $f : \{\pm 1\}^n \to \mathbb{R}$ has the Fourier expansion

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x).$$

Plancherel's identity says

$$\langle f, g \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S).$$

In particular, for Boolean f, we have Parseval's identity

$$1 = \langle f, f \rangle = \sum_{S \subseteq [n]} \hat{f}(S)^2.$$

Approximating one Fourier coefficient. We will use the following fact from the previous lecture.

Lemma 1. *For any $S \subseteq [n]$, one can estimate $\hat{f}(S)$ to within additive error γ with confidence at least $1 - \delta$ using*

$$O\left(\frac{1}{\gamma^2} \log \frac{1}{\delta}\right)$$

random examples.

3 The low-degree algorithm

We now define the notion of Fourier concentration.

Definition 2. *Let $f : \{\pm 1\}^n \to \mathbb{R}$. We say that f has $\alpha(\varepsilon, n)$-*Fourier concentration *if*

$$\sum_{\substack{S \subseteq [n] \\ |S| > \alpha(\varepsilon, n)}} \hat{f}(S)^2 \leq \varepsilon.$$

For Boolean-valued f, this is a quantitative way to say that almost all Fourier mass lies on coefficients of degree at most $\alpha(\varepsilon, n)$.

Observation 3. *If f depends on at most k variables, then for every S with $|S| > k$,*

$$\hat{f}(S) = 0.$$

Hence such an f has k-Fourier concentration.

This suggests the following learning algorithm.

Low-degree algorithm. Given parameters:

- degree bound d,
- accuracy parameter τ,
- confidence parameter δ,

do the following:

- Take
$$m = O\left(\frac{n^d}{\tau} \log \frac{n^d}{\delta}\right)$$
random examples.
- For every $S \subseteq [n]$ with $|S| \leq d$, estimate $\hat{f}(S)$ by a value C_S.
- Define
$$h(x) = \sum_{|S| \leq d} C_S \chi_S(x).$$
- Output $\operatorname{sign}(h(x))$ as the hypothesis.

The first stage of the analysis was proved last time.

Theorem 4. *If f has $d = \alpha(\varepsilon, n)$-Fourier concentration, then with probability at least $1 - \delta$, the function h output by the low-degree algorithm satisfies*

$$\mathbf{E}_x[(f(x) - h(x))^2] \leq \varepsilon + \tau.$$

To turn this into a classifier, we need to relate squared error to classification error.

Theorem 5. *Let $f : \{\pm 1\}^n \to \{\pm 1\}$ and let $h : \{\pm 1\}^n \to \mathbb{R}$. Then*

$$\Pr_x[f(x) \neq \operatorname{sign}(h(x))] \leq \mathbf{E}_x[(f(x) - h(x))^2].$$

Proof. We compare the two quantities pointwise. For each x,

$$\mathbf{1}_{\{f(x)\neq \mathrm{sign}(h(x))\}} \leq (f(x) - h(x))^2.$$

Indeed, if $f(x) = \mathrm{sign}(h(x))$, then the indicator is 0, so the inequality is trivial.

If instead $f(x) \neq \mathrm{sign}(h(x))$, then since $f(x) \in \{\pm 1\}$ and $h(x)$ has the opposite sign, $h(x)$ is at distance at least 1 from $f(x)$. Hence

$$(f(x) - h(x))^2 \geq 1$$

while

$$\mathbf{1}_{\{f(x)\neq \mathrm{sign}(h(x))\}} = 1.$$

Averaging over x proves the theorem. □

Combining the two stages, we get:

Theorem 6. *Suppose a concept class $\mathcal{C}$ has Fourier concentration*

$$d = \alpha(\varepsilon, n).$$

Then there is a uniform-distribution learning algorithm for $\mathcal{C}$ using

$$q = O\left(\frac{n^d}{\varepsilon} \log \frac{n^d}{\delta}\right)$$

random examples such that, with probability at least $1 - \delta$, the output hypothesis h' satisfies

$$\Pr[h'(x) \neq f(x)] \leq 2\varepsilon.$$

Proof. Run the low-degree algorithm with $\tau = \varepsilon$. By the previous theorem, with probability at least $1 - \delta$, we obtain h such that

$$\mathbf{E}[(f - h)^2] \leq \varepsilon + \tau = 2\varepsilon.$$

Now output $h' = \mathrm{sign}(h)$. By the classification-error bound,

$$\Pr[h'(x) \neq f(x)] \leq \mathbf{E}[(f - h)^2] \leq 2\varepsilon.$$

□

Applications

Bounded-depth decision trees

Let T be a decision tree of depth k. For each leaf ℓ, let $f_\ell(x)$ be the indicator that x reaches leaf ℓ, and let $\mathrm{val}(\ell) \in \{\pm 1\}$ be the label of the leaf. Then

$$f(x) = \sum_{\ell \text{ leaf}} f_\ell(x)\, \mathrm{val}(\ell).$$

Hence by linearity of the Fourier transform,

$$\hat{f}(S) = \sum_{\ell \text{ leaf}} \mathrm{val}(\ell)\, \hat{f}_\ell(S).$$

Each f_ℓ depends only on the variables queried along the path to ℓ, so if the tree has depth k, then

$$\hat{f}_\ell(S) = 0 \qquad \text{for all } |S| > k.$$

Therefore

$$\hat{f}(S) = 0 \qquad \text{for all } |S| > k.$$

So bounded-depth decision trees have low-degree Fourier concentration.

Constant-depth circuits

Definition 7. *A* Boolean circuit *is a directed acyclic graph whose gates are drawn from*

$$\wedge,\ \vee,\ \neg,\ 0,\ 1,\ x_1, \ldots, x_n.$$

There are strong lower bounds showing that parity cannot be computed by polynomial-size constant-depth circuits. The useful consequence for us is the following Fourier concentration theorem.

Theorem 8 (Håstad, Linial, Mansour, Nisan)**.** *If f is computable by a size-s, depth-d Boolean circuit, then for every $\alpha > 0$,*

$$\sum_{|S|>t} \hat{f}(S)^2 \leq \alpha$$

where

$$t = O\left(\log\left(\frac{s}{\alpha}\right)^{d-1}\right).$$

Thus, when $s = \mathsf{poly}(n)$ and d is constant, we get low-degree Fourier concentration. Plugging this into the low-degree algorithm yields a quasipolynomial-time sample bound.

"Lemons → lemonade." The negative result that constant-depth circuits cannot compute parity efficiently turns into a positive structural statement: such circuits must have Fourier mass concentrated on low degrees, and therefore they are learnable by the low-degree algorithm.

Learning halfspaces

Definition 9. *A* halfspace *is a function of the form*

$$h(x) = \operatorname{sign}(w \cdot x - \theta).$$

Theorem 10. *Let h be a halfspace over $\{\pm 1\}^n$. Then h has Fourier concentration*

$$\alpha(\varepsilon) = O\left(\frac{1}{\varepsilon^2}\right).$$

That is,

$$\sum_{|S|\geq c/\varepsilon^2} \hat{h}(S)^2 \leq \varepsilon$$

for a suitable constant c.

Hence the low-degree algorithm learns halfspaces under the uniform distribution using

$$n^{O(1/\varepsilon^2)}$$

random examples.

4 Noise sensitivity

We now introduce a useful way to prove Fourier concentration bounds.

Definition 11. *Fix $\varepsilon \in (0, 1/2)$. The* noise operator N_ε *acts on $x \in \{\pm 1\}^n$ by independently flipping each coordinate with probability ε.*

Definition 12. *The* noise sensitivity *of $f : \{\pm 1\}^n \to \{\pm 1\}$ is*

$$NS_\varepsilon(f) = \Pr_x[f(x) \neq f(N_\varepsilon(x))].$$

Examples

Example 13. *If $f(x) = x_1$, then*

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

since the value changes exactly when the first bit is flipped.

Example 14. *Let $f(x) = x_1 \wedge x_2 \wedge \cdots \wedge x_\ell$.*
Then

$$NS_\varepsilon(f) = \Pr[f(x) = 1 \text{ and } f(N_\varepsilon(x)) = -1] + \Pr[f(x) = -1 \text{ and } f(N_\varepsilon(x)) = 1].$$

The first term is

$$\Pr[x_1 = \cdots = x_\ell = 1] \cdot \Pr[\text{at least one of these bits flips}] = \frac{1}{2^\ell}\big(1 - (1-\varepsilon)^\ell\big).$$

The second term is

$$\Pr[\text{not all } x_i = 1] \cdot \Pr[N_\varepsilon(x)_1 = \cdots = N_\varepsilon(x)_\ell = 1].$$

For any fixed x, each coordinate becomes 1 after noise with probability $1/2$, independently, so this equals

$$\left(1 - \frac{1}{2^\ell}\right) \cdot \frac{1}{2^\ell}.$$

Thus,

$$NS_\varepsilon(f) = \frac{1}{2^\ell}\big(1 - (1-\varepsilon)^\ell\big) + \left(1 - \frac{1}{2^\ell}\right)\frac{1}{2^\ell}.$$

Example 15. *If*

$$f(x) = \mathrm{Maj}(x_1, \ldots, x_n) = \mathrm{sign}(x_1 + \cdots + x_n),$$

then

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

Remark. A high-level intuition for majority is the following. The sum $x_1 + \cdots + x_n$ behaves like a random walk and is typically of size about $\sqrt{n}$. Applying noise flips about εn bits, producing a displacement of order $\sqrt{\varepsilon n}$. The sign changes only if this displacement is large enough to cross 0, which happens with probability on the order of $\sqrt{\varepsilon}$.

In fact, the same upper bound holds for every halfspace.

Theorem 16 (Peres)**.** *For every halfspace h,*

$$NS_\varepsilon(h) \le 8.8\sqrt{\varepsilon}.$$

Example 17. *If*

$$f(x) = x_1 x_2 \cdots x_k,$$

then

$$NS_\varepsilon(f) = \Pr[\text{an odd number of bits among } 1, \ldots, k \text{ are flipped}] = \frac{1 - (1-2\varepsilon)^k}{2}.$$

In particular, for large k, this can be close to $1/2$.

Noise sensitivity and Fourier concentration

The key theorem is that noise sensitivity can be written exactly in terms of the Fourier coefficients.

Theorem 18. *For every Boolean function $f : \{\pm 1\}^n \to \{\pm 1\}$,*

$$NS_\varepsilon(f) = \frac{1}{2} - \frac{1}{2} \sum_{S \subseteq [n]} (1-2\varepsilon)^{|S|} \hat{f}(S)^2.$$

Proof. Omitted. □

This immediately yields a bound on the high-degree Fourier mass.

Theorem 19. *For every Boolean function $f : \{\pm 1\}^n \to \{\pm 1\}$ and every $\gamma \in (0, 1/2)$,*

$$\sum_{|S| \geq 1/\gamma} \hat{f}(S)^2 \leq 2.32\, NS_\gamma(f).$$

Proof. Using the previous theorem,

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

Restricting to sets of size at least $1/\gamma$, we get

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

For such S,

$$(1-2\gamma)^{|S|} \leq (1-2\gamma)^{1/\gamma} \leq e^{-2},$$

so

$$1 - (1-2\gamma)^{|S|} \geq 1 - e^{-2}.$$

Therefore,

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

which implies

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

□

Applying this to halfspaces and using Peres' theorem gives the Fourier concentration bound from earlier.

Corollary 20. *If h is a halfspace, then*

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

Proof. Choose $\gamma = \Theta(\varepsilon^2)$. Then

$$NS_\gamma(h) \leq 8.8\sqrt{\gamma} = O(\varepsilon).$$

By the previous theorem,

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

Since $1/\gamma = O(1/\varepsilon^2)$, this gives the desired concentration bound. □