Threshold Functions and Noise Sensitivity (lecture notes)

Course: Analysis of Boolean Functions, Autumn 2025, University of Chicago
Instructor: William Hoza (williamhoza@uchicago.edu)

Definition 0.1 (Linear threshold functions)**.** A *linear threshold function* (LTF) is a function $f\colon \{\pm 1\}^n \to \{\pm 1\}$ of the form $f(x) = \operatorname{sign}(w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n)$, where $w_0, w_1, \ldots, w_n \in \mathbb{R}$.

For concreteness, we use the convention $\operatorname{sign}(0) = +1$. Alternatively, by replacing w_0 with $w_0 + \varepsilon$ for some tiny ε, we can ensure that $\operatorname{sign}(0)$ never comes up. In these notes, we will use Fourier analysis to prove some Fourier concentration results, learnability results, and circuit lower bounds for classes of functions related to LTFs.

1 Fourier concentration of linear threshold functions

LTFs are unate, so they have total influence at most $\sqrt{n}$. This immediately implies that LTFs are ε-concentrated at degree up to $\sqrt{n}/\varepsilon$. We will prove a much stronger degree bound of $O(1/\varepsilon^2)$ using the notion of *noise sensitivity*.

Definition 1.1 (Noise sensitivity)**.** Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. Sample $x \in \{\pm 1\}^n$ uniformly at random, and sample

$$y_i = \begin{cases} -x_i & \text{with prob } \delta \\ x_i & \text{with prob } 1 - \delta \end{cases}$$

independently for each i. Then we define $\mathrm{NS}_\delta[f] = \Pr[f(x) \neq f(y)]$.

Noise sensitivity is just a reparameterized version of noise stability. Specifically, from the definitions, one can double check that

$$\operatorname{Stab}_\rho[f] = 1 - 2\,\mathrm{NS}_{(1-\rho)/2}[f]$$
$$\text{and } \mathrm{NS}_\delta[f] = \frac{1}{2} - \frac{1}{2}\operatorname{Stab}_{1-2\delta}[f].$$

We will show that if $\mathrm{I}[f]$ is small *and* we have some mild closure properties, then f has low noise sensitivity. Then we will show that if f has low noise sensitivity, then it is concentrated at low degree.

1.1 Bounded total influence $\implies$ bounded noise sensitivity

Theorem 1.2. *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. Let $m \leq n$, and let $\mathcal{C}$ be the class of functions $f'\colon \{\pm 1\}^m \to \{\pm 1\}$ one can get by starting with f and then identifying and negating variables. Assume that every function in $\mathcal{C}$ has total influence at most I. Then $\mathrm{NS}_{1/m}[f] \leq I/m$.*

Proof. As a thought experiment, sample $w \in \{\pm 1\}^m$, $\pi\colon [n] \to [m]$, $z \in \{\pm 1\}^n$, and $j \sim [m]$ uniformly at random. Define

$$x = z \circ w^\pi$$
$$y = z \circ (w^{\oplus j})^\pi.$$

Then x is distributed uniformly over $\{\pm1\}^n$, and y can be constructed from x by flipping each bit independently with probability $1/m$. Therefore, if we define $g_{z,\pi}(w) = f(z \circ w^\pi)$, then

$$\begin{aligned}\mathrm{NS}_{1/m}[f] = \Pr_{w,\pi,z,j}[f(z \circ w^\pi) \neq f(z \circ (w^{\oplus j})^\pi] &= \mathop{\mathbb{E}}_{z,\pi}\left[\frac{1}{m} \cdot \sum_{j=1}^{m} \Pr_w[g_{z,\pi}(w) \neq g_{z,\pi}(w^{\oplus j})]\right] \\ &= \frac{1}{m} \mathop{\mathbb{E}}_{z,\pi}[\mathrm{I}[g_{z,\pi}]] \\ &\leq I/m. \qquad \square\end{aligned}$$

Corollary 1.3 (Peres's Theorem)**.** *If f is an LTF, then* $\mathrm{NS}_\delta[f] \leq O(\sqrt{\delta})$.

Proof. Let $m = \lfloor 1/\delta \rfloor$. Then $\delta \leq 1/m$, so $\mathrm{NS}_\delta[f] \leq \mathrm{NS}_{1/m}[f]$. If we identify/negate variables of f, we get another LTF on m variables, which is unate, hence it has total influence at most $\sqrt{m}$. Therefore, $\mathrm{NS}_{1/m}[f] \leq \sqrt{m}/m = 1/\sqrt{m} \leq O(\sqrt{\delta})$. □

Note that the bound goes to 0 *uniformly* as $\delta \to 0$, i.e., there is no dependence on n.

Corollary 1.4. *If f has the form $\mathsf{ANY}_s \circ \mathsf{LTF}$, then* $\mathrm{NS}_\delta[f] \leq O(s \cdot \sqrt{\delta})$.

Proof. Union bound. □

For example, an "s-facet polytope" is a function of the form $\mathsf{AND}_s \circ \mathsf{LTF}$.

1.2 Bounded noise sensitivity $\implies$ concentration at low degrees

Theorem 1.5. *Let $f\colon \{\pm1\}^n \to \{\pm1\}$. Then f is $O(\mathrm{NS}_{1/k}[f])$-concentrated on degree up to k.*

Proof.

$$\begin{aligned}2\,\mathrm{NS}_{1/k}[f] = 1 - \mathrm{Stab}_{1-2/k}[f] = \mathop{\mathbb{E}}_{S \sim \mathcal{S}_f}[1 - (1-2/k)^{|S|}] &\geq (1-(1-2/k)^k) \cdot \Pr_{S \sim \mathcal{S}_f}[|S| \geq k] \\ &\geq (1-e^{-2}) \cdot W^{\geq k}[f]. \qquad \square\end{aligned}$$

Consequently, LTFs are ε-concentrated on degree up to $O(1/\varepsilon^2)$. More generally, we have shown that (roughly speaking) a total influence bound of the form $n \cdot \varepsilon(n)$ implies $O(\varepsilon(k))$-concentration up to degree k.

2 Chow's theorem

An "$\mathsf{LTF} \circ \mathsf{LTF}$ circuit" is a function of the form $f(x) = g(h_1(x), \ldots, h_m(x))$, where $g, h_1, \ldots, h_m$ are all LTFs. The *size* of the circuit is $m+1$, the total number of LTFs. This is a simple type of *neural network*. How powerful are $\mathsf{LTF} \circ \mathsf{LTF}$ circuits?

Every function $f\colon \{\pm1\}^n \to \{\pm1\}$ can be computed by an $\mathsf{LTF} \circ \mathsf{LTF}$ circuit of exponential size, because we can put f in conjunctive / disjunctive normal form. Note that AND and OR are both special cases of LTFs. What about polynomial-size $\mathsf{LTF} \circ \mathsf{LTF}$ circuits?

We will prove that there are (many) functions that cannot be computed by polynomial-size $\mathsf{LTF} \circ \mathsf{LTF}$ circuits. The key is Chow's theorem, which says that an LTF is fully specified by its Fourier coefficients of degree at most 1.

Theorem 2.1 (Chow's theorem)**.** *Let $f\colon \{\pm1\}^n \to \{\pm1\}$ be an LTF and let $g\colon \{\pm1\}^n \to \mathbb{R}$. Assume $\widehat{f}(S) = \widehat{g}(S)$ for every $S \subseteq [n]$ such that $|S| \leq 1$. Then $f \equiv g$.*

Proof. Since f is an LTF, there is some degree-1 function $\ell\colon \{\pm 1\}^n \to \mathbb{R}$ such that for every $x \in \{\pm 1\}^n$, we have $\ell(x) \neq 0$ and $f(x) = \operatorname{sign}(\ell(x))$. Then

$$\begin{aligned}\mathop{\mathbb{E}}_x[|\ell(x)|] = \mathop{\mathbb{E}}_x[f(x)\cdot \ell(x)] = \sum_{S\subseteq[n]} \widehat{f}(S)\cdot\widehat{\ell}(S) = \sum_{|S|\le 1} \widehat{f}(S)\cdot\widehat{\ell}(S) &= \sum_{|S|\le 1} \widehat{g}(S)\cdot\widehat{\ell}(S)\\ &= \sum_{S\subseteq[n]} \widehat{g}(S)\cdot\widehat{\ell}(S)\\ &= \mathop{\mathbb{E}}_x[g(x)\cdot\ell(x)]\end{aligned}$$

Since $\ell(x)$ is never zero, this implies that $g(x) = \operatorname{sign}(\ell(x))$ for every x. □

Corollary 2.2. *The number of distinct functions $f\colon \{\pm 1\}^n \to \{\pm 1\}$ that can be computed by LTFs is $2^{O(n^2)}$.*

Proof. For any Boolean function $f\colon \{\pm 1\}^n \to \{\pm 1\}$, every Fourier coefficient is an integer multiple of 2^{-n} between -1 and 1. Hence there are $O(2^n)$ possibilities for each of the $n+1$ Fourier coefficients of degree ≤ 1, so we get a bound of $O(2^n)^{n+1}$. □

Corollary 2.3. *The number of distinct functions $f\colon \{\pm 1\}^n \to \{\pm 1\}$ that can be computed by size-m $\mathsf{LTF}\circ\mathsf{LTF}$ circuits is $2^{O(n^2 m)}$.*

In contrast, the total number of functions $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is 2^{2^n}, which is much larger than $2^{O(n^2 m)}$ if $m = \operatorname{poly}(n)$.

3 Polynomial threshold functions

The proof in the previous section was dissatisfying, because it was nonconstructive. It is a major open problem to prove that an "explicit" Boolean function cannot be computed by polynomial-size $\mathsf{LTF}\circ\mathsf{LTF}$ circuits.[1] One possible candidate hard function is the *inner product mod* 2 function. If n is even, then we define $\mathsf{IP}_n\colon \{\pm 1\}^n \to \{\pm 1\}$ by the formula

$$\mathsf{IP}_n(x_1,\dots,x_{n/2},y_1,\dots,y_{n/2}) = \prod_{i=1}^{n/2} \max(x_i,y_i).$$

(The name comes from the fact that if we encode inputs and outputs using $\{0,1\}$, then the formula would become $\mathsf{IP}_n(x,y) = \sum_i x_i y_i \bmod 2$. But we will encode bits using ± 1 for convenience.) We will prove that IP_n cannot be computed by sparse *polynomial threshold functions*, which are an interesting special case of $\mathsf{LTF}\circ\mathsf{LTF}$ circuits.

Definition 3.1 (Polynomial threshold functions)**.** A *polynomial threshold function (PTF) of sparsity m* is a function $f\colon \{\pm 1\}^n \to \{\pm 1\}$ of the form $f(x) = \operatorname{sign}(p(x))$, where p is a multilinear polynomial with at most m monomials.

Proposition 3.2. *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. If f can be computed by a PTF of sparsity m, then f can also be computed by an $\mathsf{LTF}\circ\mathsf{LTF}$ circuit of size at most $mn+1$.*

Proof. If $x \in \{0,1\}^n$, then

$$\begin{aligned}\bigoplus_{i=1}^n x_i = \sum_{i=1}^n x_i \bmod 2 &= \sum_{t=1}^n (-1)^{t+1}\cdot 1\left[\sum_{i=1}^n x_i \ge t\right]\\ &= \sum_{t=1}^n (-1)^{t+1}\cdot\left(\frac{1}{2}+\frac{1}{2}\cdot\operatorname{sign}\left(0.5 - t + \sum_{i=1}^n x_i\right)\right).\end{aligned}$$

[1] For example, it is an open problem to prove that there exists a function $f \in \mathsf{NP}$ that cannot be computed by polynomial-size $\mathsf{LTF}\circ\mathsf{LTF}$ circuits.

Therefore, if $x \in \{\pm 1\}^n$, then

$$\begin{aligned}
\prod_{i \in S} x_i &= 1 - 2 \cdot \bigoplus_{i=1}^{n} \left(\frac{1}{2} - \frac{1}{2} x_i \right) \\
&= 1 - 2 \cdot \sum_{t=1}^{n} (-1)^{t+1} \cdot \left(\frac{1}{2} + \frac{1}{2} \cdot \mathrm{sign} \left(0.5 - t + \sum_{i=1}^{n} \left(\frac{1}{2} - \frac{1}{2} x_i \right) \right) \right) \\
&= \text{a constant} + \text{a linear combination of } n \text{ LTFs.}
\end{aligned}$$

Therefore, if $f(x) = \mathrm{sign}(\sum_{i=1}^{m} c_i \prod_{j \in S_i} x_j)$, then

$$\begin{aligned}
f(x) &= \mathrm{sign} \left(\sum_{i=1}^{m} c_i \cdot (\text{a constant} + \text{a linear combination of at most } n \text{ LTFs}) \right) \\
&= \mathrm{sign} \left(\text{a constant} + \text{a linear combination of at most } mn \text{ LTFs} \right). \qquad \square
\end{aligned}$$

Lemma 3.3. *Let f be a PTF, $f(x) = \mathrm{sign}(p(x))$. Assume f is not constant. Then*

$$\sum_{\substack{S \subseteq [n] \\ \widehat{p}(S) \neq 0}} |\widehat{f}(S)| \geq 1.$$

Proof. Let S_* maximize $|\widehat{p}(S_*)|$. Then

$$|\widehat{p}(S_*)| = \left| \mathop{\mathbb{E}}_{x}[p(x) \cdot \chi_{S_*}(x)] \right| \leq \mathop{\mathbb{E}}_{x}[|p(x)|] = \mathop{\mathbb{E}}_{x}[p(x) \cdot f(x)] = \sum_{\widehat{p}(S) \neq 0} \widehat{p}(S) \cdot \widehat{f}(S) \leq |\widehat{p}(S_*)| \cdot \sum_{\widehat{p}(S) \neq 0} |\widehat{f}(S)|.$$

Since f is not constant, we can divide both sides by $|\widehat{p}(S_*)|$. $\square$

Theorem 3.4. *If n is even, then every PTF computing IP_n has sparsity at least $2^{n/2}$.*

Proof. On your homework (Exercise 3), you prove that $|\widehat{\mathsf{IP}}_n(S)| = 2^{-n/2}$ for every $S \subseteq [n]$. $\square$