1 Peres’s Theorem

Recall the definition of noise sensitivity:

**Definition 1.**

\[
\text{NS}_\epsilon(f) = \Pr_{x,y \sim \mathcal{N}_\epsilon(x)}[f(x) \neq f(y)]
\]

In previous lectures, we have seen \(\text{NS}_\epsilon(\text{PAR})_k = \frac{1}{2} (1 - (1 - 2\epsilon)^k)\), and for any \(f : \{-1,1\}^n \rightarrow \{-1,1\}\) such that \(\text{NS}_\epsilon(f) = \frac{1}{2} \left(1 - \sum_{S \subset [n]} (1 - 2\epsilon)^{|S|} \hat{f}(S)^2\right)\).

**Lemma 2.** Fix any \(f : \{-1,1\}^n \rightarrow \{-1,1\}\), any noise rate \(0 < \gamma < \frac{1}{2}\). Then

\[
\sum_{|S| > \frac{1}{\gamma}} \hat{f}(S)^2 \leq 2.32 \text{NS}_\gamma(f)
\]

**Proof.** As previous lectures shown,

\[
2\text{NS}_\gamma(f) = 1 - \sum_{S \subset [n]} (1 - 2\gamma)^{|S|} \hat{f}(S)^2
\]

By Parsevals Theorem,

\[
2\text{NS}_\gamma(f) = 1 - \sum_{S \subset [n]} (1 - 2\gamma)^{|S|} \hat{f}(S)^2
\]

\[
= \sum_{S \subset [n]} [1 - (1 - 2\gamma)^{|S|}] \hat{f}(S)^2
\]

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

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

\[\square\]
Corollary 3. Let \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \), let \( \beta : [0, \frac{1}{2}] \rightarrow [0, \frac{1}{2}] \) be a function such that
\[
\text{NS}_\gamma(f) \leq \beta(\gamma),
\]
then for \( m = \frac{1}{\beta^{-1}(\epsilon/2.32)} \),
\[
\sum_{|S| > m} \hat{f}(S)^2 \leq \epsilon.
\]

Proof. We have
\[
\sum_{|S| > m} \hat{f}(S)^2 = \sum_{1/\beta^{-1}(\epsilon/2.32)} \hat{f}(S)^2 \leq 2.32 \text{NS}_{\beta^{-1}(\epsilon/2.32)}(f) \leq 2.32(\beta^{-1}(\epsilon/2.32)) = \epsilon.
\]

□

Theorem 4 (Peres’s Theorem). Let \( f(x) = \text{sign}(w \cdot x - \theta) \), \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \).

Then \( \text{NS}_\epsilon(f) \leq 2\sqrt{\epsilon} \)

Let \( \beta^{-1}(x) = (x/2)^2 \). It is easy to show that the corresponding \( \beta \) function satisfies the condition of Corollary 3. Thus, \( \beta^{-1}(\epsilon/2.32) \leq (\epsilon/5)^2 \).

Corollary 5. Every \( n \)-variable LTF \( f \) satisfies following condition:
\[
\sum_{|S| > 25/\epsilon^2} \hat{f}(S)^2 \leq \epsilon
\]

Let \( h_1, h_2, \ldots, h_K \) be \( K \) halfspaces over \( \{-1, 1\}^n \). Let \( g : \{-1, 1\}^K \rightarrow \{-1, 1\} \) be any function and \( f(x) = g(h_1(x), h_2(x), \ldots, h_K(x)) \). We have following theorem:

Theorem 6.
\[
\text{NS}_\epsilon(f) \leq 2K\sqrt{\epsilon}
\]

Proof. By definition of noise sensitivity,
\[
\text{NS}_\epsilon(f) = \Pr_{x, y \sim \mathbb{N}_\epsilon(x)}[f(x) \neq f(y)] = \Pr_{x, y \sim \mathbb{N}_\epsilon(x)}[g(h_1(x), \ldots, h_K(x)) \neq g(h_1(y), \ldots, g_K(y))] \leq 2K\sqrt{\epsilon}
\]

□

Theorem 7. For \( f = g(h_1, \ldots, h_K) \), where \( g \) is an arbitrary \( K \)-variable boolean function and \( h_1, \ldots, h_K \) are arbitrary LTFs.
\[
\sum_{|S| > 25K^2/\epsilon^2} \hat{f}(S)^2 \leq \epsilon
\]
Remark 1. It is interesting to study whether \( 25K^2/\epsilon^2 \) can be improved.

For \( f = \text{MAJ}(x_1, x_2, \ldots, x_n) \),

\[
\sum_{|S| > K} \hat{f}(S)^2 = \Theta(1/\sqrt{K}).
\]

Thus, the constant 2 in the “\( \frac{1}{\sqrt{2}} \)” can not be improved.

If \( g \) is \( \text{PAR} \) over \( K \) variables and \( h_1 = x_1, \ldots, h_K = x_K \), \( f(x) = \text{PAR}(x_1, \ldots, x_K) \) has 0 Fourier weight at levels lower than \( K \). So the constant 2 of “\( K^2 \)” can not improve to constant smaller than 1.

For \( f = h_1 \land h_2 \land \cdots \land h_K \), \( \sum_{|S| > O(\log K/\epsilon^2)} \hat{f}(S)^2 \leq \epsilon \) holds for Gaussian space.

1.1 Proof of Peres’s Theorem

In following proof, we will assume \( \epsilon = \frac{1}{t} \) for some integer \( t \geq 1 \), and prove \( \text{NS}_\epsilon(f) \leq \sqrt{\epsilon} \) for \( f(x) = \text{sign}(w \cdot x - \theta) \). It is an official homework that following proof implies Peres’s Theorem.

Lemma 8. If \( \epsilon = \frac{1}{t} \) for some integer \( t \geq 1 \), \( \text{NS}_\epsilon(f) \leq \sqrt{\epsilon} \) for \( f(x) = \text{sign}(w \cdot x - \theta) \).

Proof. We have

\[
\text{NS}_\epsilon(f) = \Pr_{x, y \sim N_\epsilon(x)} [f(x) \neq f(y)] = \Pr_{b, \alpha, \epsilon, j} [f(x) \neq f(y)]
\]

We consider \( x \) and \( y = N_\epsilon(x) \). Let \( b_1, b_2, \ldots, b_n \) be sampled independent uniformly from \( \{1, \ldots, t\} \). Let \( b_1, b_2, \ldots, b_n \) be sampled uniformly from \( \{1, \ldots, t\} \). Let \( z = z_1, z_2, \ldots, z_{t-1}, -z_j, z_{j+1}, \ldots, z_t \) and \( y_i = \alpha_i \cdot z_i \). It is easy to verify that \( x_i \) is equivalent to be independently sampled from \( \{1, \ldots, t\} \) and \( y = N_\epsilon(x) \).

For any fixed \( b, \alpha \), the value of \( f(x) \) is determined as a function of \( z \in \{-1, 1\}^t \).

Let \( h_{\alpha, b}(z) = f(x) \), we have

\[
\text{NS}_\epsilon(f) = \Pr_{b, \alpha, z, j} [f(x) \neq f(y)] = \mathbf{E}_{\alpha, b, z, j} [1_{h_{\alpha, b}(z) \neq h_{\alpha, b}(z^{\ominus j})}] = \mathbf{E}_{\alpha, b} \left[ \frac{1}{t} \sum_{j=1}^t \mathbf{E}_z [h_{\alpha, b}(z) \neq h_{\alpha, b}(z^{\ominus j})] \right]
\]

Since \( h_{\alpha, b}(z) \neq h_{\alpha, b}(z^{\ominus j}) = \text{Inf}(h_{\alpha, b}) \) and \( \sum_{j=1}^t \mathbf{E}_z [h_{\alpha, b}(z) \neq h_{\alpha, b}(z^{\ominus j})] = \text{Inf}(h_{\alpha, b}) \), we have \( \text{NS}_\epsilon(f) = \mathbf{E}_{\alpha, b} \left[ \frac{1}{t} \text{Inf}(h_{\alpha, b}(z)) \right] \).

For \( \alpha, b \), \( h_{\alpha, b}(z) \) is a \( t \)-variable LTF such that

\[
h_{\alpha, b}(z) = \text{sign}(w_1 \alpha_1 z_{b_1} + \cdots + w_n \alpha_n z_{b_n} - \theta),
\]
which is a LTF over $z_1, z_2, \ldots, z_t$. Thus, for $\alpha, b$, $\inf(t - \text{variable LTF}) \leq \sqrt{t}$. Finally, we have
\[
\NS_{\epsilon}(f) \leq \avg_{\alpha, b}(\sqrt{t}/t) = \sqrt{\epsilon}
\]

\[\square\]

\section{Learning beyond Fourier concentration}

In this section, we will learn boolean functions without Fourier concentration approach. One possible approach is to look at low degree Fourier coefficients, but do more than just “weight them”. We will us this approach to learn LTFs and random DNFs.

\subsection{Learning LTFs}

\textbf{Definition 9.} The “Chow parameter” of a $n$-variable LTF are the degree-0 and degree-1 Fourier coefficients:
\[
\hat{f}(0) = \hat{f}(\emptyset), \hat{f}(1), \ldots, \hat{f}(n)
\]

\textbf{Theorem 10.} Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be an LTF $f(x) = \sign(w \cdot x - \theta)$. Let $g : \{-1, 1\}^n \to \{-1, 1\}$. Suppose $\hat{g}(i) = \hat{h}(i)$ for all $0 \leq i \leq n$, then $g(x) = f(x)$ for all $x \in \{-1, 1\}^n$.

\textit{Proof.} Let $f(x) = \sign(w_0 + w_1 x_1 + \cdots + w_n x_n)$. Without loss of generality, we have $w_0 + w_1 x_1 + \cdots w_n x_n > 0$ for any $x \in \{-1, 1\}^n$. Thus, we have
\[
\E[(w_0 + w_1 x_1 + \cdots + w_n x_n)(f(x) - g(x))] = \sum_{j=0}^{n} w_j (\hat{f}(j) - \hat{g}(j)) = 0
\]
by Plancherels Theorem.

On the other hand, if $f(x) - g(x) < 0$, we know $f(x) = -1$ and $g(x) = 1$. By the definition of $f$, $w_0 + w_1 x_1 + \cdots + w_n x_n < 0$ implies $(w_0 + w_1 x_1 + \cdots w_n x_n < 0)(f(x) - g(x)) > 0$. If $f(x) - g(x) > 0$, we know $f(x) = 1$ and $g(x) = -1$. Again $w_0 + w_1 x_1 + \cdots w_n x_n > 0$ implies $(w_0 + w_1 x_1 + \cdots + w_n x_n < 0)(f(x) - g(x)) > 0$. This means $(w_0 + w_1 x_1 + \cdots + w_n x_n)(f(x) - g(x))$ is always non-negative. Thus, for any $x \in \{-1, 1\}^n$, $f(x) = g(x)$.

\textbf{Remark 2.} It is possible to efficiently construct $\epsilon$-approximation of LTF given accurate estimation of Chow parameter of an LTF.
2.2 Learning random DNFs

Definition 11. Let $f = T_1 \lor T_2 \lor \cdots \lor T_n$ be a $n$-term DNF. Each term $T_i$ is generated in following way:

1. Pick a set $S_i \subset [n]$ such that $|S_i| = \log n$.
2. For each $j \in S_i$, let $l_{ij} = x_j$ with probability $1/2$ and $l_{ij} = \bar{x}_j$ with probability $1/2$.
3. $T_i = \land_{j \in S_i} l_{ij}$.

With high probability, a random DNF generated in this way can be reconstructed by only looking at $\hat{f}(S)$ for $|S| \leq 2$. 