Analysis of Boolean Functions

CS 395T – Spring 2020 – Dr. Anna Gal

Yangxinyu Xie

Department of Computer Science

University of Texas at Austin

The textbook used in this class is [O'D14].
Last updated June 9, 2020

1 Motivation

1.1 Boolean Functions ❖

Usually, a **Boolean function** f maps a length-n binary vector, or *string*, into a single binary value, or *bit*. The most basic examples seen in computer science are

$$\begin{array}{ll} \vee & \texttt{OR}(x_1, x_2, ..., x_n) \\ \wedge & \texttt{AND}(x_1, x_2, ..., x_n) \\ \oplus & \texttt{PARITY}(x_1, x_2, ..., x_n) \end{array}$$

In graph theory, given an undirected graph $G = (V, E), |V| = m$ where the ordering of the adjacency matrix is pre-specified, we can represent the graph G as a binary string of length $n = \binom{m}{2}$. Hence, for instance, we can view the algorithm for the clique problem as a Boolean function. That is, the algorithm that determines if a k clique exists in a graph with m vertices can be written as

$$\texttt{CLIQUE}_{k,m} : \{0,1\}^{\binom{m}{2}} \to \{0,1\} \tag{1.1.1}$$

Generalising this idea, we can also define a Boolean function $g : \{0,1\}^{2n} \to \{0,1\}^{2n}$ that maps a string to another string. For example, binary integer multiplication takes in two n-bit strings and outputs a $2n$-bit string.
In fact, any discrete yes-no problem can be viewed as searching for a family of Boolean functions that maps a fixed-length string to a bit, since a language $L \in \{0,1\}^*$ can be divided into disjoint *slices*

$$L = \bigcup_n L_n \text{ where } L_n \in \{0,1\}^n \tag{1.1.2}$$

1.2 Truth Table and Normal Forms ❖

A Boolean function $f : \{0,1\}^n \to \{0,1\}$ can also be viewed as a truth table. Specifically, given a pre-specified ordering, we will have a table with 2^n indices, each of which denotes a n-bit string in $\{0,1\}^n$. The entires of the table are simply binary numbers. For example, a function $f : \{0,1\}^3 \to \{0,1\}$ may be represented as below

000	0
100	0
010	0
001	1
011	1
101	1
110	1
111	1

Any Boolean function can be expressed as a disjunctive normal form. The DNF of the above example is

$$(\bar{x}_1 \wedge \bar{x}_2 \wedge x_3) \vee (\bar{x}_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge \bar{x}_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge \bar{x}_3) \vee (x_1 \wedge x_2 \wedge x_3) \tag{1.2.1}$$

2 Fourier Expansion

2.1 DNF to Polynomial ❖

Notation 2.1. Given $\alpha \in \{0,1\}$, we write

$$x_i^\alpha = \begin{cases} x_i & \text{if } \alpha = 1 \\ \bar{x}_i & \text{if } \alpha = 0 \end{cases} \tag{2.1.1}$$

Then given a string $a_1 a_2 ... a_n$, we have that $x_1^{a_1} x_2^{a_2} ... x_n^{a_n} = 1$ if and only if x and a are equal.

Hence, we obtain

$$x_1^{a_1} x_2^{a_2} ... x_n^{a_n} = \prod_{i:a_i=1} x_i \prod_{i:a_i=0} \bar{x}_i = \prod_{i:a_i=1} x_i \prod_{i:a_i=0} (1 - x_i) \tag{2.1.2}$$

The second equality follows from that x_i is binary. Any disjunctive normal form can be written as the following

$$f(x) = \bigvee_{a \in \{0,1\}^n, f(a)=1} \prod_{i:a_i=1} x_i \prod_{i:a_i=0} (1 - x_i) \tag{2.1.3}$$

Since there is only one element $a \in \{0,1\}^n$ that equals x, we have

$$f(x) = \sum_{a \in \{0,1\}^n, f(a)=1} \prod_{i:a_i=1} x_i \prod_{i:a_i=0} (1 - x_i) = \sum_{a \in \{0,1\}^n} f(a) \prod_{i:a_i=1} x_i \prod_{i:a_i=0} (1 - x_i) \tag{2.1.4}$$

Hence, we obtain a polynomial representation of $f(x)$.

2.2 Fourier Expansion Theorem ❖

To obtain a simpler representation of the polynomial 2.1.4, we first adapt the following replacement

$$\begin{array}{ccc} 1 & \to & -1 \\ 0 & \to & 1 \end{array}$$

which gives us

$$\frac{1 + x_i}{2} = \begin{cases} 0 & \text{if } x_i = -1 \\ 1 & \text{if } x_i = 1 \end{cases} \qquad \frac{1 - x_i}{2} = \begin{cases} 1 & \text{if } x_i = -1 \\ 0 & \text{if } x_i = 1 \end{cases} \tag{2.2.1}$$

which allows us to write the counterpart of 2.1.4 as an function $f : \{-1,1\}^n \to \{0,1\}$

$$f(x) = \sum_{a \in \{-1,1\}^n} f(a) \prod_{i:a_i=-1} \frac{1 - x_i}{2} \prod_{i:a_i=1} \frac{1 + x_i}{2} \tag{2.2.2}$$

We can rewrite f as a function from $\{-1,1\}^n$ to $\{-1,1\}$ by some algebra,

$$f(x) = 1 - 2 \sum_{a \in \{-1,1\}^n} \frac{1 - f(a)}{2} \prod_{i:a_i=-1} \frac{1 - x_i}{2} \prod_{i:a_i=1} \frac{1 + x_i}{2} \tag{2.2.3}$$

By expanding the polynomial above 2.2.3, we can rewrite f as a sum of monomials

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i \tag{2.2.4}$$

where $\prod_{i \in S} x_i$ is in fact the parity function and is defined to be 1 if $S = \emptyset$ and $\hat{f}(S)$ is the coefficient of the monomial $\prod_{i \in S} x_i$.

Remark 2.2. Since $x_i \in \{-1, 1\}$, we see that $\prod_{i \in S} x_i = -1$ if there is an odd number of -1 or `TRUE` values, $\prod_{i \in S} x_i$ is the parity function.

Theorem 2.3 (**Fourier Expansion Theorem**). *Every function $f : \{-1, 1\}^n \to \mathbb{R}$ can be uniquely expressed as a multilinear polynomial,*

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) x^S \tag{2.2.5}$$

where $x^S = \prod_{i \in S} x_i$. This expression is called the **Fourier expansion** *of f, and the real number $\hat{f}(S)$ is called the* **Fourier coefficient** *of f on S. Collectively, the coefficients are called the* **Fourier spectrum** *of f.*

We can prove the uniqueness of the multilinear polynomial by showing the linear independence of the 2^n different x^S. In fact, we can show that the set of x_S forms an orthonormal basis of the space of the functions $f : \{-1, 1\}^n \to \mathbb{R}$.

2.3 The Orthonormal Basis ❖

Definition 2.4. We define the **inner product** $\langle \cdot, \cdot \rangle$ on pairs of function $f, g : \{-1, 1\}^n \to \mathbb{R}$ by

$$\langle f, g \rangle = 2^{-n} \sum_{x \in \{-1,1\}^n} f(x) g(x) \tag{2.3.1}$$

From a probabilistic perspective, if we let x be *uniformly distributed* in $\{-1, 1\}^n$, we have

$$\langle f, g \rangle = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x) g(x)] \tag{2.3.2}$$

Lemma 2.5.

$$\langle f, f \rangle = 2^{-n} \sum_{x \in \{-1,1\}^n} f(x) f(x) = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x)^2] \tag{2.3.3}$$

Theorem 2.6. *The 2^n different parity functions $x^S : \{-1, 1\}^n \to \{-1, 1\}$ form an orthonormal basis for the vector space V spanned by functions $f : \{-1, 1\}^n \to \mathbb{R}$. In other words, for any two sets $S, T \in \{-1, 1\}^n$,*

$$\langle x^S, x^T \rangle = \begin{cases} 1 & \text{if } S = T \\ 0 & \text{if } S \neq T \end{cases} \tag{2.3.4}$$

A direct corollary from this theorem shows the uniqueness of the Fourier expansion.

Corollary 2.7. *The 2^n different parity functions $x^S : \{-1, 1\}^n \to \{-1, 1\}$ are linear independent and hence the scalars $\hat{f}(S)$ of the linear combination of x^S*

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) x^S \tag{2.3.5}$$

is unique.

Lemma 2.8. *For any two sets $S, T \in \{-1, 1\}^n$, we have*

$$x^S x^T = \begin{cases} 1 & \text{if } S = T \\ x^{S \Delta T} & \text{if } S \neq T \end{cases} \tag{2.3.6}$$

where $S \Delta T = (S \setminus T) \cup (T \setminus S)$ is the symmetric difference of S and T.

Proof. If $S = T$, we then have $x^S x^T = (x^S)^2 = 1$. If $S \neq T$, then

$$x^S x^T = \prod_{i \in S} x_i \prod_{i \in T} x_i = \prod_{i \in S \cap T} x_i^2 \prod_{i \in S \Delta T} x_i = \prod_{i \in S \Delta T} x_i = x^{S \Delta T} \tag{2.3.7}$$

□

Lemma 2.9.

$$\mathbb{E}[x^S] = \begin{cases} 1 & \text{if } S = \emptyset \\ 0 & \text{if } S \neq \emptyset \end{cases} \tag{2.3.8}$$

Proof. $x^\emptyset = 1$ by definition. By linearity of expectation, we have

$$\mathbb{E}[x^S] = \mathbb{E}[\prod_{i \in S} x_i] = \prod_{i \in S} \mathbb{E}[x_i] \tag{2.3.9}$$

Since x is uniformly distributed in $\{-1, 1\}^n$, x_i must be uniformly distributed in $\{-1, 1\}$. Hence, $\mathbb{E}[x_i] = 0$ and concludes the proof. □

proof of Theorem 2.6. Note that $\langle x^S, x^T \rangle = \mathbb{E}_{x \sim \{-1,1\}^n}[x^S x^T]$. If $S = T$, we have the expectation is just the constant 1

$$\mathbb{E}_{x \sim \{-1,1\}^n}[x^S x^T] = 1 \tag{2.3.10}$$

If $S \neq T$, we then have

$$\mathbb{E}_{x \sim \{-1,1\}^n}[x^S x^T] = \mathbb{E}_{x \sim \{-1,1\}^n}[x^{S \Delta T}] = 0 \tag{2.3.11}$$

since $S \Delta T$ is not empty. □

3 Inner Product

Some results directly following the inner products are shown in this section.

Notation 3.1. We use the notation χ_S and x^S interchangeably to denote $\prod_{i\in S} x_i$.

Theorem 3.2. *For $f : \{-1,1\}^n \to \mathbb{R}$ and $S \subseteq [n]$, we have*

$$\langle f, \chi_S \rangle = \hat{f}(S) \tag{3.0.1}$$

Proof.

$$\begin{aligned}
\langle f, \chi_S \rangle &= \langle \sum_{T\subseteq[n]} \hat{f}(T)\chi_T, \chi_S \rangle \\
&= 2^{-n} \sum_{x\in\{-1,1\}^n} \sum_{T\subseteq[n]} \hat{f}(T)\chi_T(x)\chi_S(x) \quad \text{by Fourier expansion} \\
&= \sum_{T\subseteq[n]} \hat{f}(T)[2^{-n} \sum_{x\in\{-1,1\}^n} \chi_T(x)\chi_S(x)] \quad \text{by linearity of expectation} \\
&= \sum_{T\subseteq[n]} \hat{f}(T)\langle \chi_T, \chi_S \rangle \\
&= \hat{f}(S) \quad \text{by theorem 2.6}
\end{aligned} \tag{3.0.2}$$

□

Theorem 3.3 (Parseval's Theorem). *For any $f : \{-1,1\}^n \to \mathbb{R}$, we have*

$$\langle f, f \rangle = \sum_{S\subseteq[n]} \hat{f}(S)^2 = \mathbb{E}_{x\sim\{-1,1\}^n}[f(x)^2] \tag{3.0.3}$$

This theorem can be seen as a direct corollary from the following theorem.

Theorem 3.4 (Plancherel's Theorem). *For any $f, g : \{-1,1\}^n \to \mathbb{R}$, we have*

$$\langle f, g \rangle = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S) = \mathbb{E}_{x\sim\{-1,1\}^n}[f(x)g(x)] \tag{3.0.4}$$

Proof.

$$\begin{aligned}
\langle f, g \rangle &= \langle \sum_{S\subseteq[n]} \hat{f}(S)\chi_S, \sum_{T\subseteq[n]} \hat{g}(T)\chi_T \rangle \\
&= 2^{-n} \sum_{x\in\{-1,1\}^n} \sum_{S\subseteq[n]} \hat{f}(S)\chi_S(x) \sum_{T\subseteq[n]} \hat{g}(T)\chi_T(x) \quad \text{by Fourier expansion} \\
&= \sum_{S\subseteq[n]} \sum_{T\subseteq[n]} \hat{f}(S)\hat{g}(T)[2^{-n} \sum_{x\in\{-1,1\}^n} \chi_S(x)\chi_T(x)] \quad \text{by linearity of expectation} \\
&= \sum_{S\subseteq[n]} \sum_{T\subseteq[n]} \hat{f}(S)\hat{g}(T)\langle \chi_S, \chi_T \rangle \\
&= \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S) \quad \text{by theorem 2.6}
\end{aligned} \tag{3.0.5}$$

□

Note that $\langle f, g \rangle$ can be seen as a measure of similarity between the functions f and g. To see this, we first give the definition of distance.

Definition 3.5. Given two Boolean functions $f, g : \{-1,1\}^n \to \{-1,1\}$, we define the **distance** between f and g to be

$$\text{dist}(f,g) = \mathbb{P}_{x\sim\{-1,1\}^n}[f(x) \neq g(x)] \tag{3.0.6}$$

Since x is chosen uniformly, we have equivalently $\text{dist}(f,g) = 2^{-n}$(Hamming distance between f and g). By the definition of inner product 2.4, we have that

$$\langle f, g\rangle = 2^{-n} \sum_{x\in\{-1,1\}^n} f(x)g(x) \tag{3.0.7}$$

Because $f(x)g(x) = 1$ if $f(x) = g(x)$ and $f(x)g(x) = -1$ otherwise, we have the following proposition.

Proposition 3.6. *For any $f, g : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\langle f, g\rangle = 2^{-n} \sum_{x\in\{-1,1\}^n} f(x)g(x) = \mathbb{P}_{x\sim\{-1,1\}^n}[f(x) = g(x)] - \mathbb{P}_{x\sim\{-1,1\}^n}[f(x) \neq g(x)] = 1 - 2\text{dist}(f,g) \tag{3.0.8}$$

The following corollary is thus a direct result of this proposition.

Corollary 3.7. *For two sets $S, T, S \neq T$, we have*

$$\text{dist}(\chi_S, \chi_T) = \frac{1}{2} \tag{3.0.9}$$

Proof. By lemma 2.8, we have $\langle \chi_S, \chi_T\rangle = 0$. Hence, $1 - 2\text{dist}(f,g) = 0$ by proposition 3.6. □

4 BLR Test

4.1 Almost Linear Functions and BLR Test ❖

Lemma 4.1. *For any $f : \{0,1\}^n \to \{0,1\}$, we have the following are equivalent*

(i) $f(x+y) = f(x) + f(y)$

(ii) $f(x) = \sum_{i \in A} x_i$ *for some* $A \subseteq [n]$

Proof. (i) $\Rightarrow$ (ii): For any $x \in \{0,1\}^n$, we can write

$$x = x_1 e_1 + x_2 e_2 + ... + x_n e_n \tag{4.1.1}$$

Thus by (i), we have

$$f(x) = f(x_1 e_1) + f(x_2 e_2) + ... + f(x_n e_n) \tag{4.1.2}$$

Notice that by (i), the additive identity must exist. In other words, $f(0) = 0$. Hence, if $x_i = 0$, we have $f(0) = 0 = x_i f(e_i)$. If $x_i = 1$, then $f(x_i e_i) = f(e_i) = x_i f(e_i)$. Therefore,

$$f(x) = x_1 f(e_1) + x_2 f(e_2) + ... + x_n f(e_n) = \sum_{i \in A} x_i \quad \text{where } A = \{i | f(e_i) = 1\} \tag{4.1.3}$$

(ii) $\Rightarrow$ (i):

$$f(x) + f(y) = \sum_{i \in A} x_i + \sum_{i \in A} y_i = \sum_{i \in A} (x_i + y_i) = f(x+y) \tag{4.1.4}$$

□

Definition 4.2. A function $f : \{0,1\}^n \to \{0,1\}$ is **linear** if one of the following holds:

(i) $f(x+y) = f(x) + f(y)$

(ii) $f(x) = \sum_{i \in A} x_i$ for some $A \subseteq [n]$

Definition 4.3. For any Boolean functions f, g, we say that f and g are **ϵ-close** if $\text{dist}(f,g) < \epsilon$.

One idea of *property testing* has the following scenario. Suppose we have a "black-box" which contains an unknown function $f : \{0,1\}^n \to \{0,1\}$ and we want to verify if this function is linear. Specifically, given an input $x \in \{0,1\}^n$, we can query $f(x)$ quickly but verifying all 2^n possible values is too expensive. Hence, we aim to test the linearity property approximately by the following scheme:

Algorithm 4.4 (BLR Test). Given query access to a function $f : \{0,1\}^n \to \{0,1\}$:

1. Sample $x, y \in \{0,1\}^n$ uniformly and independently.
2. Query $f(x), f(y)$ and $f(x+y)$.
3. `Accept` if $f(x) + f(y) = f(x+y)$.

Switching to $f : \{-1,1\}^n \to \{0,1\}$, we have the analogous algorithm

Algorithm 4.5 (BLR Test). Given query access to a function $f : \{-1,1\}^n \to \{-1,1\}$:

1. Sample $x, y \in \{-1,1\}^n$ uniformly and independently.
2. Query $f(x), f(y)$ and $f(x \circ y)$ where $x \circ y = (x_1 y_1, x_2, y_2, ..., x_n y_n)$.
3. `Accept` if $f(x) f(y) f(x \circ y) = 1$.

Note that linear functions are character functions in $\{-1,1\}$.

Lemma 4.6.

$$\mathbb{P}[\text{BLR Test } \texttt{Accept}] = \frac{1}{2} + \frac{1}{2}\sum_{S\subseteq[n]} \hat{f}(S)^3 \tag{4.1.5}$$

Proof. Note that

$$\frac{1}{2} + \frac{1}{2}f(x)f(y)f(x\circ y) = \begin{cases} 1 & \text{if BLR accepts} \\ 0 & \text{otherwise} \end{cases} \tag{4.1.6}$$

Hence, we have

$$\begin{aligned} \mathbb{P}[\text{BLR Test } \texttt{Accept}] &= \mathbb{P}[\frac{1}{2} + \frac{1}{2}f(x)f(y)f(x\circ y)] \\ &= \frac{1}{2} + \frac{1}{2}\mathbb{P}[f(x)f(y)f(x\circ y)] \\ &= \frac{1}{2} + \frac{1}{2}\mathbb{E}[f(x)f(y)f(x\circ y)] \\ &= \frac{1}{2} + \frac{1}{2}\mathbb{E}[\sum_{S\in[n]}\hat{f}(S)\chi_S(x)\sum_{T\in[n]}\hat{f}(T)\chi_T(y)\sum_{U\in[n]}\hat{f}(U)\chi_U(x\circ y)] \\ &= \frac{1}{2} + \frac{1}{2}\mathbb{E}[\sum_{S\in[n]}\hat{f}(S)\sum_{T\in[n]}\hat{f}(T)\sum_{U\in[n]}\hat{f}(U)\chi_S(x)\chi_T(y)\chi_U(x\circ y)] \\ &= \frac{1}{2} + \frac{1}{2}\sum_{S\in[n]}\hat{f}(S)\sum_{T\in[n]}\hat{f}(T)\sum_{U\in[n]}\hat{f}(U)\mathbb{E}[\chi_S(x)\chi_T(y)\chi_U(x\circ y)] \end{aligned} \tag{4.1.7}$$

where

$$\begin{aligned} \mathbb{E}[\chi_S(x)\chi_T(y)\chi_U(x\circ y)] &= \mathbb{E}_{x,y\sim\{-1,1\}^n}[\prod_{i\in S}x_i\prod_{j\in T}y_i\prod_{k\in U}x_ky_k] \\ &= \mathbb{E}_{x,y\sim\{-1,1\}^n}[\prod_{i\in S\Delta U}x_i\prod_{j\in T\Delta U}y_i] \\ &= \mathbb{E}_{x\sim\{-1,1\}^n}[\prod_{i\in S\Delta U}x_i]\mathbb{E}_{y\sim\{-1,1\}^n}[\prod_{j\in T\Delta U}y_i] \quad x,y \text{ are independent} \\ &= \mathbb{E}[\chi_{S\Delta U}]\mathbb{E}[\chi_{T\Delta U}] \end{aligned} \tag{4.1.8}$$

By Lemma 2.9, we obtain

$$\begin{aligned} \mathbb{P}[\text{BLR Test } \texttt{Accept}] &= \frac{1}{2} + \frac{1}{2}\sum_{S\in[n]}\hat{f}(S)\sum_{T\in[n]}\hat{f}(T)\sum_{U\in[n]}\hat{f}(U)\mathbb{E}[\chi_{S\Delta U}]\mathbb{E}[\chi_{T\Delta U}] \\ &= \frac{1}{2} + \frac{1}{2}\sum_{S\in[n]}\hat{f}(S)\hat{f}(S)\hat{f}(S) \\ &= \frac{1}{2} + \frac{1}{2}\sum_{S\in[n]}\hat{f}(S)^3 \end{aligned} \tag{4.1.9}$$

□

Theorem 4.7. *If* $\mathbb{P}[\text{BLR Test } \texttt{Accept}] > 1-\epsilon$*, then* f *is* ϵ*-close to some linear function* $g : \{0,1\}^n \to \{0,1\}$*.*

Proof. By Lemma 4.6, we have

$$\mathbb{P}[\text{BLR Test } \texttt{Accept}] = \frac{1}{2} + \frac{1}{2}\sum_{S\in[n]}\hat{f}(S)^3 > 1-\epsilon \tag{4.1.10}$$

which gives us

$$1 - 2\epsilon < \sum_{S\in[n]} \hat{f}(S)^3 \leq \max_{T\in[n]} f(T) \sum_{S\in[n]} \hat{f}(S)^2 = \max_{T\in[n]} f(T) \tag{4.1.11}$$

By Theorem 3.2, we have

$$1 - 2\epsilon < \max_{T\in[n]} f(T) = \max_{T\in[n]} \langle f, \chi_T \rangle \tag{4.1.12}$$

By Proposition 3.6, we obtain

$$1 - 2\epsilon < \max_{T\in[n]} (1 - 2\text{dist}(f, \chi_T)) \tag{4.1.13}$$

Hence, we have that f is close to some linear function χ_T. □

Theorem 4.8. *For any $f : \{0,1\}^n \to \{0,1\}$, we have the following are equivalent*

(i) $f(x+y) = f(x) + f(y)$ for most pairs of $x, y \in \{0,1\}^n$.

(ii) $f(x) = \sum_{i\in A} x_i$ for some $A \subseteq [n]$ for most inputs $x \in \{0,1\}^n$.

proof sketch. (i) ⇒ (ii): Since BLR Test accepts if $f(x+y) = f(x) + f(y)$ for most pairs of $x, y \in \{0,1\}^n$, theorem 4.7 implies (ii).
(ii) ⇒ (i): $f(x) + f(y) = \sum_{i\in A} x_i + \sum_{i\in y} x_i = \sum_{i\in A}(x_i + y_i)$ for most inputs $x, y \in \{0,1\}^n$. □

Algorithm 4.9. How to use BLR Test

1. Repeat BLR Test for $O(1/\epsilon)$ times
2. If BLR Test fails for some input x, y, `Reject`.
3. If BLR Test did not fail, then f is ϵ-close to some linear function.

5 Local Decoding and List Decoding

Error-correcting codes are used to ensure reliable transmission of information over noisy channels as well as to ensure reliable storage of information on a medium that may be partially corrupted over time.
A **code** of length N is simply a subset $C \in A^N$, where A is a finite set. Elements of C are called **codewords**. A binary code is one with alphabet $A = \{0, 1\}$. An r-query locally decodable code C encodes k-bit messages m in such a way that one can probabilistically recover any bit $m(i)$ of the message by querying only r bits of the (possibly corrupted) codeword $C(m)$, where r can be as small as 2.

Definition 5.1. A q-ary code $C : \mathbb{F}_q^n \to \mathbb{F}_q^N$ is said to be (r, δ, ϵ)**-locally decodable** if there exists a randomised decoding algorithm $\mathcal{A}$ such that for all $m \in \mathbb{F}_q^n$ and all $w \in \mathbb{F}_q^N$ such that $\text{dist}(C(m), w) \leq \delta$:

1. For every index $i \in [n]$
$$\mathbb{P}[\mathcal{A}(w, i) = m_i] \geq 1 - \epsilon, \tag{5.0.1}$$
where the probability is taken over the random coin tosses of the algorithm $\mathcal{A}$.

2. $\mathcal{A}$ makes at most r queries to w.

A similar definition considers recovering any codeword position as opposed to any message bit.

Definition 5.2. A q-ary code $C : \mathbb{F}_q^n \to \mathbb{F}_q^N$ is said to be (r, δ, ϵ)**-locally correctable** if there exists a randomised decoding algorithm $\mathcal{A}$ such that for all $m \in \mathbb{F}_q^n$ and all $w \in \mathbb{F}_q^N$ such that $\text{dist}(C(m), w) \leq \delta$:

1. For every index $j \in [N]$
$$\mathbb{P}[\mathcal{A}(w, j) = C(m)_j] \geq 1 - \epsilon, \tag{5.0.2}$$
where the probability is taken over the random coin tosses of the algorithm $\mathcal{A}$.

2. $\mathcal{A}$ makes at most r queries to w.

5.1 Hadamard Code ❖

The classical Hadamard code encoding n-bit messages to 2^n-bit codewords is a *2-query locally decodable code*, and also a *2-query locally correctable code*.
Every codeword in the Hadamard code corresponds to one of 2^n subsets S of $[n]$: the codewords of the Hadamard code are the truth tables of χ_S, for $S \subseteq [n]$. The positions of the codewords are indexed by strings $x \in \{0, 1\}^n$. For message $S \subseteq [n]$, the position of the codeword χ_S indexed by x is simply $\chi_S(x)$ (that is the parity of the bits of x corresponding to the subset S.) Let w be a possibly corrupted encoding of S. (That is, instead of the truth table of χ_S, w is the truth table of some other function, not too far from χ_S.) First we assume that the corrupted words w are within relative distance $\delta <= 1/4$ from SOME correct codeword.
First we see that the Hadamard code is locally correctable. Given $x \in \{0, 1\}^n$, and w, the decoder picks random $y \in \{0, 1\}^n$, queries $w(y)$ and $w(x + y)$, and outputs the parity of these two bits. If w differs from the correct encoding of S in at most δ fraction of coordinates, then, by union bound, with probability at least $1 - 2\delta$ both of the decoder's queries will go to uncorrupted locations. In such case, the decoder correctly recovers $\chi_S(x)$.
Next we consider local decoding, that is recovering individual message bits. Given an index $i \in [n]$ and w, the Hadamard decoder picks a string $y \in \{0, 1\}^n$ uniformly at random and outputs the parity of the two coordinates of w corresponding to $w(y)$ and $w(y + e_i)$. As before, if w differs from the correct encoding of S in at most δ fraction of coordinates, then, by union bound, with probability at least $1 - 2\delta$ both of the decoder's queries will go to uncorrupted locations. In such case, the decoder correctly recovers the value $\chi_S(e_i)$, which tells us wheter or not $i \in S$.
In fact, Hadamard code achieves optimal length for 2-query locally decodable codes.

Theorem 5.3. *If there exists an $(2, \delta, \epsilon)$-locally decodable code C encoding n-bit messages to N-bit codewords; then*
$$N \geq 2^{\Omega((1/2-\epsilon)^4 \delta^2 n)} \tag{5.1.1}$$

The proof uses Quantum information argument and can be found in [Y+12]. □

For 3-query locally decodable codes, Yekhanin [Yek08] constructs a code with sub-exponential length of size $\exp(\exp(O(\log n/(\log\log n))))$.
Notice that if the distance between a corrupted codeword and some codeword is at most half of the minimum distance, which is 1/4 for Hadamard code, then we have *unique decoding*.
On the other hand, for Hadamard codes, since we have for any $f:\{-1,1\}^n \to \{-1,1\}$,

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 = 1 \tag{5.1.2}$$

there are only a constant number of large coefficients of the Fourier expansion of f. This means that even if a corrupted codeword (corresponding to some Boolean function f) is at distance greater than 1/4, there will only be a constant number of possible codewords. To see this, let $\text{dist}(f,\chi_S) \le 1/2-\epsilon$, we have

$$\hat{f}(S) = \langle f, \chi_S\rangle = 1 - 2\text{dist}(f,\chi_S) \ge 2\epsilon \tag{5.1.3}$$

Because there are at most $1/(4\epsilon^2)$ coefficients of f whose magnitude is greater than 2ϵ, there are at most $1/(4\epsilon^2)$ number of possible parity functions within distance $1/2-\epsilon$ from f.

Remark 5.4. For every small $\delta > 0$, the number of subsets S such that $|\hat{f}(S)| \ge \delta$ is at most $1/\delta^2$.

Therefore, we can use **list decoding** to recover a list of possible codewords within the given distance to the corrupted received word.

Definition 5.5. [GRS12] Given $0 \le \rho \le 1, l \ge 1$, a code $C \in \mathbb{F}_q^N$ is (ρ, l)-list decodable if for every received word $y \in \mathbb{F}_q^N$,

$$|\{x \in C | \text{dist}(y,x) \le \rho\}| \le l \tag{5.1.4}$$

Theorem 5.6 (Goldreich-Levin Theorem)**.** *There is a probabilistic algorithm such that given $f:\{0,1\}^n \to \{0,1\}$ and $\gamma,\delta > 0$, it runs in time poly$(n, 1/\gamma)\log(1/\delta)$ and with probability at least $1-\delta$ outputs a list of subsets $L = \{S_1, ..., S_k\}$ such that*

1. *if $|\hat{f}(S)| \ge \gamma$, then $S \in L$.*
2. *if $S \in L$, then $|\hat{f}(S)| \ge \gamma/2$*

This theorem says that the Hadamard code is efficiently list decodable.

5.2 Goldreich-Levin Algorithm ❖

The Goldreich-Levin Algorithm uses a divide and conquer scheme to estimate the Fourier weight of f on various collections of sets.

Algorithm 5.7. Goldreich-Levin Algorithm

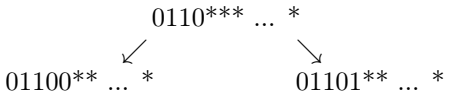
1. Given some integer $k \le n$ and some subset $S \subseteq [k]$, the **bucket** $B_{k,S}$ is the collection of sets such that

$$B_{k,S} := \{S \cup T : T \subseteq [n] \setminus [k]\}. \tag{5.2.1}$$

   We start with $k = 0$ and $S = \emptyset$.

2. `Repeat` the following until each bucket has only 1 set.

   - Select any bucket with 2^m sets where $m = n - k$.
   - Split the bucket $B_{k,S}$ into $B_{k+1,S}$ and $B_{k+1,S\cup\{k+1\}}$. In other words, we split the bucket $B_{k,S}$ by fixing the $k+1$th bit.



- Estimate the weight of each bucket by estimating

$$\sum_{U \in B} \hat{f}(U)^2 \tag{5.2.2}$$

- discard any bucket B if its weight estimate $\leq \gamma^2/2$

3. `Output` the list L of remaining buckets.

To see the correctness of the algorithm, we assume that we can efficiently estimate the weight of each bucket and the estimates are accurate within $\gamma^2/4$. Therefore,

1. If $\hat{f}(U) \geq \gamma$, we have $\hat{f}(U)^2 \geq \gamma^2 \geq \gamma^2/2 + \gamma^2/4$ and U will never be discarded.
2. If $\hat{f}(U) < \gamma/2$, we have $\hat{f}(U)^2 < \gamma^2/4 = \gamma^2/2 - \gamma^2/4$ and U will not remain in the end.

As for running time, by Remark 5.4, we see that there are at most $4/\gamma^2$ buckets in the end. Because each of the surviving bucket will experience at most n splits, there are at most $4n/\gamma^2$ repetitions of the main loop.

Lemma 5.8 (Chernoff's Bound)**.** *Let $X := \frac{\sum_{i\in[n]} X_i}{n}$ where $X_i, i \in [n]$ are independent, identically distributed, real valued random variables in [a, b]. Then,*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \epsilon] \leq 2\exp(-\frac{2n\epsilon^2}{b-a}) \tag{5.2.3}$$

Lemma 5.9. *Given $f : \{-1,1\}^n \to \{-1,1\}, \gamma, \delta > 0$ and $S \subseteq [n]$, we can estimate $\hat{f}(S)$ within error $\pm\gamma$ with probability $1-\delta$ using $O(\frac{1}{\gamma^2}\log\frac{1}{\delta})$ queries.*

Proof. Since $\hat{f}(S) = \mathbb{E}_{x\sim\{-1,1\}^n}[f(x)\chi_S(x)]$, by Chernoff's Bound, we only need $O(\frac{1}{\gamma^2}\log\frac{1}{\delta})$ queries to achieve error $\pm\gamma$ with probability $1-\delta$. □

Given a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, let $x = y||z$, where $a||b$ is a concatenation of a and b. Using Fourier expansion, we have

$$\begin{aligned} f(x) = \sum_{U\subseteq[n]} \hat{f}(U)x^U &= \sum_{S\subseteq[k], T\subseteq[n]\setminus[k]} \hat{f}(S\cup T)y^S z^T \\ &= \sum_{S\in[k]} y^S \Big[\sum_{T\subseteq[n]\setminus[k]} \hat{f}(S\cup T)z^T\Big] \end{aligned} \tag{5.2.4}$$

For fixed $z \in \{-1,1\}^{n-k}$, denote by f_{*z} the subfunction from $\{-1,1\}^k$ to $\{-1,1\}$. Then looking at the Fourier expansion of $f(x)$ above, we see that

$$\hat{f}_{*z}(S) := \sum_{T\subseteq[n]\setminus[k]} \hat{f}(S\cup T)z^T \tag{5.2.5}$$

Claim 5.10.

$$\sum_{U\in B_{k,S}} \hat{f}(U)^2 = \sum_{T\subseteq[n]\setminus[k]} \hat{f}(S\cup T)^2 = \mathbb{E}_{z\sim\{-1,1\}^{n-k}}[\hat{f}_{*z}(S)^2] \tag{5.2.6}$$

Proof. The first equality follows from the definition of the bucket.
To show the second equality, we let $F(z) : \{-1,1\}^{n-k} \to \mathbb{R}$ be defined as $F(z) := \hat{f}_{*z}(S)$. The intuition is that we suppose S is fixed and view z as a variable; hence, $F(z) = \hat{f}_{*z}(S)$ is then a function of z. Since

$$F(z) = \hat{f}_{*z}(S) = \sum_{T\subseteq[n]\setminus[k]} \hat{f}(S\cup T)z^T \tag{5.2.7}$$

we obtain

$$\hat{F}(z) = \hat{f}(S \cup T) \tag{5.2.8}$$

Hence,

$$\mathbb{E}_{z\sim\{-1,1\}^{n-k}}[\hat{f}_{*z}(S)^2] = \mathbb{E}_{z\sim\{-1,1\}^{n-k}}[F(z)^2] = \sum_{T\subseteq[n]\setminus[k]} \hat{F}(z)^2 = \sum_{T\subseteq[n]\setminus[k]} \hat{f}(S\cup T)^2 \tag{5.2.9}$$

The second equality is obtained by Parseval's Theorem 3.3. □

By theorem 3.2, we have

$$\hat{f}_{*z}(S) = \langle f_{*z}, \chi_S\rangle = \mathbb{E}_{y\sim\{-1,1\}^k}[f(y||z)\chi_S(y)] \tag{5.2.10}$$

which implies

$$\begin{aligned}\mathbb{E}_{z\sim\{-1,1\}^{n-k}}[\hat{f}_{*z}(S)^2] &= \mathbb{E}_{z\sim\{-1,1\}^{n-k}}\left[\left(\mathbb{E}_{y\sim\{-1,1\}^k}[f(y||z)\chi_S(y)]\right)^2\right] \\ &= \mathbb{E}_{z\sim\{-1,1\}^{n-k}}\mathbb{E}_{y\sim\{-1,1\}^k}\mathbb{E}_{y'\sim\{-1,1\}^k}[f(y||z)\chi_S(y)f(y'||z)\chi_S(y')]\end{aligned} \tag{5.2.11}$$

where y and y' are independent. Similar to Lemma 5.9, we only need to make $O(\frac{1}{\gamma^4}\log\frac{1}{\delta})$ queries to estimate $\sum_{U\in B}\hat{f}(U)^2$ with error $\gamma^2/4$ and probability $1-\delta$. Consequently, we have proved the Goldreich-Levin Theorem 5.6.

6 Learning Theory

To learn a character function χ_S exactly, we only need n queries, each being e_i, given that each query outputs a correct answer.
Suppose, on the other hand, we have query access to a function $f : \{-1,1\}^n \to \{-1,1\}$ that is ϵ-close to some character function χ_S. As discussed in section 5.1, for each $i \in [n]$, if we pick a random y, we have $\chi_S(e_i) = f(y) \cdot f(e_i \circ y)$ with probability at least $1 - 2\epsilon = 1/2 + \gamma$. Using Chernoff's Bound 5.8, if we repeat $O(\frac{\log(n/\delta)}{4\gamma^2})$ times and take the majority of $f(y) \cdot f(e_i \circ y)$, then we have

$$\mathbb{P}[\text{majority incorrect}] \leq \frac{\delta}{n} \tag{6.0.1}$$

which means we can get $\chi_S(e_i)$ correctly for all e_i with probability at least $1 - n \cdot \frac{\delta}{n} = 1 - \delta$.

6.1 "PAC" Learning ❖

Notation 6.1. We use $\mathscr{C}$ to denote a **concept class**, which is a collection of functions f.

Definition 6.2. In the model of "PAC" (Probably Approximately Correct) learning, a **learning algorithm** $\mathcal{A}$ for $\mathscr{C}$ is a randomised algorithm which has limited access (for example, a polynomial number) to an **unknown target function** $f \in \mathscr{C}$ and outputs some efficient representation of f. Typically, we have two **access models**:

- Random Examples: $\mathcal{A}$ can draw pairs $(x, f(x))$ where $x \in$ is uniformly random.
- Membership Queries: $\mathcal{A}$ can request the value of $f(x)$ where x is chosen by the algorithm $\mathcal{A}$.

We say that $\mathcal{A}$ **learns** a concept class $\mathscr{C}$ with error ϵ if it learns an $f \in \mathscr{C}$ with error ϵ. That is, with high probability, $\mathcal{A}$ outputs a **hypothesis function** $h : \{-1,1\}^n \to \{-1,1\}$ which is ϵ-close to f.

Definition 6.3. Let $\mathcal{F}$ be a family of subsets $S \subseteq [n]$. We say that the Fourier spectrum of $f : \{-1,1\}^n \to \mathbb{R}$ is **ϵ-concentrated** on $\mathcal{F}$ if

$$\sum_{S \notin \mathcal{F}} \hat{f}(S)^2 \leq \epsilon \tag{6.1.1}$$

The following lemma is a direct result from the definition.

Lemma 6.4. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be ϵ-concentrated on $\mathcal{F}$. For $g := \sum_{S \in \mathcal{F}} \hat{f}(S)\chi_S$, we have*

$$||f - g||_2^2 \leq \epsilon \tag{6.1.2}$$

Lemma 6.5. *Suppose that $f : \{-1,1\}^n \to \{-1,1\}, g : \{-1,1\}^n \to \mathbb{R}$ satisfy that $||f - g||_2^2 \leq \epsilon$. Define a Boolean function $h : \{-1,1\}^n \to \{-1,1\}$ by $h := \text{sign}(g(x))$, then*

$$\text{dist}(f, h) \leq \epsilon \tag{6.1.3}$$

Proof. For $x \in \{-1,1\}^n$ such that $f(x) \neq h(x)$, we have $(f(x) - g(x))^2 \geq 1$. Hence,

$$||f - g||_2^2 = \mathbb{E}_{x \sim \{-1,1\}^n}[(f(x) - g(x))^2] \geq \mathbb{P}_{x \sim \{-1,1\}^n}[f(x) \neq h(x)] = \text{dist}(f, h) \tag{6.1.4}$$

□

Theorem 6.6. *Suppose an algorithm $\mathcal{A}$ has random example access to the target function $f : \{-1,1\}^n \to \{-1,1\}$ and can identify a family $\mathcal{F}$ on which f is $\epsilon/2$-concentrated. Then in poly$(|\mathcal{F}|, n, 1/\epsilon)$ time, $\mathcal{A}$ can with high probability output a hypothesis function h which is ϵ-close to f.*

Proof. The main scheme consists of two steps:

1. Since $\hat{f}(S) = \mathbb{E}_{x\sim\{-1,1\}^n}[f(x)\chi_S(x)]$, we can use the mean of samples of $f(x)\chi_S(x)$ to estimate $\hat{f}(S)$. By Chernoff's Bound 5.8, for each $S \in \mathcal{F}$, we can estimate $\hat{f}(S)$ within error $\pm\frac{\sqrt{\epsilon}}{2\sqrt{|\mathcal{F}|}}$ with probability $1-\delta$ for some small $\epsilon, \delta > 0$. Let $\widetilde{f}(S)$ be the estimates of $\hat{f}(S)$.
   The number of samples needed is $O(\frac{|\mathcal{F}|}{\epsilon}\log(1/\delta))$. Take $\delta = 1/(10|\mathcal{F}|)$, we can estimate $\hat{f}(S)$ for all S within $\text{poly}(|\mathcal{F}|, n, 1/\epsilon)$ time.
2. Let $g = \sum_{S\in\mathcal{F}} \widetilde{f}(S)\chi_S$. Output $h = \text{sign}(g)$.

It remains to show that $||f-g||_2^2 \le \epsilon$. That is

$$\begin{aligned} ||f-g||_2^2 = \langle f-g, f-g\rangle &= \sum_{S\subseteq[n]} \widehat{f-g}(S)^2 \\ &= \sum_{S\in\mathcal{F}} (\hat{f}(S) - \widetilde{f}(S))^2 + \sum_{S\notin\mathcal{F}} \hat{f}(S)^2 \\ &\le \sum_{S\in\mathcal{F}} \Big(\frac{\sqrt{\epsilon}}{2\sqrt{|\mathcal{F}|}}\Big)^2 + \frac{\epsilon}{2} \\ &\le \frac{\epsilon}{4} + \frac{\epsilon}{2} < \epsilon \end{aligned} \tag{6.1.5}$$

Hence, using lemma 6.5, we conclude the proof. □

Theorem 6.7 ([KM93]). *Let $\mathscr{C}$ be a concept class such that $f \in \mathscr{C}, f : \{-1,1\}^n \to \{-1,1\}$ has its Fourier spectrum $\epsilon/4$-concentrated on at most M sets. Then $\mathscr{C}$ can be learned with error ϵ with time poly$(M, n, 1/\epsilon)$.*

Proof. Firstly, by theorem 5.6, with high probability, we can find a list L such that if $|\hat{f}(S)| \ge \frac{\epsilon}{4M}$, then $S \in L$. Secondly, by theorem 6.6, it suffice to show that f is $\epsilon/2$-concentrated on L. Let $\mathcal{M}$ be the family of at most M sets such that f is $\epsilon/4$-concentrated on $\mathcal{M}$. Thus,

$$\begin{aligned} \sum_{S\notin L} \hat{f}(S)^2 &= \sum_{S\notin L, S\in\mathcal{M}} \hat{f}(S)^2 + \sum_{S\notin L, S\notin\mathcal{M}} \hat{f}(S)^2 \\ &\le M\cdot\frac{\epsilon}{4M} + \frac{\epsilon}{4} = \frac{\epsilon}{2} \end{aligned} \tag{6.1.6}$$

□

Theorem 6.8 ([LMN89]). *Let $\mathscr{C}$ be a concept class such that $f \in \mathscr{C}, f : \{-1,1\}^n \to \{-1,1\}$ has its Fourier spectrum $\epsilon/4$-concentrated on*

$$\mathcal{M} = \{S \subseteq [n] : |S| \le d\} \tag{6.1.7}$$

Then $\mathscr{C}$ can be learned with error ϵ with time poly$(|\mathcal{M}|, n, 1/\epsilon)$ by using the random example access model.

Proof. By theorem 6.6 with $\mathcal{F} = \mathcal{M}$. Notice that

$$|\mathcal{M}| \le \sum_{i=0}^{k} \binom{n}{i} \le O(n^k) \tag{6.1.8}$$

□

7 Decision Trees

Definition 7.1. A **decision tree** is a rooted binary tree in which each internal node is labelled by a variable x_i, each outgoing edge is labeled by $+1$ or -1, and each leaf is labelled by $+1$ or -1.

Definition 7.2. We say that a given decision tree T **computes** a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ if for every input $x \in \{-1,1\}^n$, the label of the reached leaf on input x equals $f(x)$. The **size** of T is the number of leaves of T. The **depth** of T is the length of the longest path in T.

We use $\mathbf{1}_p : \{-1,1\}^n \to \{0,1\}$ to denote the **indicator function** of a path p from the root to a leaf in a decision tree T.

$$\mathbf{1}_p(x) = \begin{cases} 1 & \text{if on input } x, \text{ we reach the leaf of the path } p \\ 0 & \text{otherwise} \end{cases} \tag{7.0.1}$$

We insist that each variable x_i is queried by the path p at most once. Equivalently, let V be the set of variables queried along the path p, we have

$$\mathbf{1}_p(x) = \prod_{i \in V_1} (\frac{1+x_i}{2}) \prod_{j \in V_2} (\frac{1-x_j}{2}) \tag{7.0.2}$$

where V_1 is the set of variables along the path p whose immediate outgoing edge is labelled 1 and V_2 is the set of those labelled -1.

Let label(p) to denote the label of the leaf of p. Then for a decision tree T that computes $f : \{-1,1\}^n \to \{-1,1\}$,

$$f(x) = \sum_{\text{path } p \in T} \mathbf{1}_p(x)\text{label}(p) \tag{7.0.3}$$

Definition 7.3. The **decision tree complexity** of f, denoted by $D(f)$, is the depth of the decision tree T such that T computes f with the least possible depth. Note that $D(f) \leq n$ for all f.

The following result is a direct observation from equation 7.0.2.

Lemma 7.4. *If a path p has length d, then the indicator function has a Fourier representation as*

$$\mathbf{1}_p(x) = \sum_{S \subseteq V} \pm\frac{1}{2^d}\chi_S(x) \tag{7.0.4}$$

where V is the set of variables queried along the path p.

Theorem 7.5. *If f is computed by a decision tree of depth d, then*

1. *The degree of f is at most d.*
2. *Each coefficient is an integer multiple of $1/2^d$.*
3. *The number of non-zero Fourier coefficients is at most 4^d.*

Proof. 1 follows from equation 7.0.2.
2 is a direct result of Lemma 7.4.
3 Since each Fourier coefficient is at least $1/2^d$, by Parseval's Theorem 3.3, we have there are at most $(2^d)^2$ non-zero Fourier coefficients. We can also see this by noticing that each indicator function has at most 2^d nonzero Fourier coefficients, and there are at most 2^d leaves. □

Corollary 7.6. *A decision tree of depth d can be learned exactly with membership queries in poly(4^d)poly(n) time.*

Proof. By setting $\epsilon = 1/2^{d+1}$, we can apply Kushilevitz-Mansour algorithm (theorem 6.7). Then, since the output function is ϵ-close to the decision tree, there is a unique decoding. In other words, we can round each Fourier coefficient estimate of the output function to the nearest multiple of $1/2^d$. □

Lemma 7.7. *Size s decision trees are ϵ-close to a depth $\log(s/\epsilon)$ decision tree.*

Proof. Let T be a size s decision tree computing $f : \{-1,1\}^n \to \{-1,1\}$, we can obtain a new tree T' by cutting off all paths of length greater than $\log(s/\epsilon)$. Observe that $T(x) \neq T'(x)$ if and only if input x follows one of the longer paths.
For any given path p of length k, since each internal node has two out-coming edges, the fraction of inputs that follow path p is at most 2^{-k}. For any $k \geq \log(s/\epsilon)$, we have such probability is thus at most ϵ/s.
Since T is of size s, we have

$$\mathbb{P}[T(x) \neq T'(x)] \leq s \cdot \epsilon/s = \epsilon \tag{7.0.5}$$

□

Corollary 7.8. *A function $f : \{-1,1\}^n \to \{-1,1\}$ computable by size s decision trees are $O(\epsilon)$-concentrated on a family of size at most $4^{\log(s/\epsilon)} = (s/\epsilon)^2$.*

The following corollary is a result of the combination of corollary 7.8 and the Linial-Mansour-Nisan algorithm (theorem 6.8).

Corollary 7.9. *A polynomial sized decision tree can be learned exactly with membership queries in polynomial time.*

8 Fourier Norms

Definition 8.1. The **Fourier (or spectral) p-norm** of a function $f : \{-1,1\}^n \to \mathbb{R}$ is

$$\hat{||}f\hat{||}_p := \Big(\sum_{S \subseteq [n]} |\hat{f}(S)|^p \Big)^{1/p} \tag{8.0.1}$$

The L_1 Fourier norm is just the sum of the absolute value of the Fourier coefficients:

$$\hat{||}f\hat{||}_1 := \sum_{S \subseteq [n]} |\hat{f}(S)| \tag{8.0.2}$$

It is easy to check the following properties.

Lemma 8.2. *For L_1 Fourier norm,*

- $\hat{||}f+g\hat{||}_1 \leq \hat{||}f\hat{||}_1 + \hat{||}g\hat{||}_1$
- $\hat{||}c \cdot f\hat{||}_1 \leq |c| \cdot \hat{||}f\hat{||}_1$ *for some constant c.*

Example 8.3. $\hat{||}\mathtt{AND}_n\hat{||}_1 = 1$.

By Parseval's Theorem 3.3, we have the L_2 Fourier norm is the same as the L_2 norm

$$\hat{||}f\hat{||}_2 := \Big(\sum_{S \subseteq [n]} |\hat{f}(S)|^2 \Big)^{1/2} = \langle f, f \rangle^{1/2} = ||f||_2 \tag{8.0.3}$$

Lemma 8.4. *For some $\epsilon > 0$, the function $f : \{-1,1\}^n \to \{-1,1\}$ is ϵ-concentrated on the set*

$$\mathcal{F} = \{S \subseteq [n] || \hat{f}(S) \geq \frac{\epsilon}{\hat{||}f\hat{||}_1}\} \tag{8.0.4}$$

Proof.

$$\begin{aligned} \sum_{S \notin \mathcal{F}} \hat{f}(S)^2 &\leq \max_{S \notin \mathcal{F}} |\hat{f}(S)| \cdot \sum_{S \notin \mathcal{F}} |\hat{f}(S)| \\ &\leq \max_{S \notin \mathcal{F}} |\hat{f}(S)| \cdot \sum_{S \subseteq [n]} |\hat{f}(S)| \\ &\leq \frac{\epsilon}{\hat{||}f\hat{||}_1} \cdot \hat{||}f\hat{||}_1 = \epsilon \end{aligned} \tag{8.0.5}$$

□

Remark 8.5. By Parseval's Theorem 3.3, we have that $|\mathcal{F}| \leq (\hat{||}f\hat{||}_1/\epsilon)^2$.

9 DNF Formulas

Definition 9.1. A **DNF(disjunctive normal form) formula** over Boolean input $x = (x_1, ..., x_n)$ is defined as a logical `OR` of **terms**, each of which is a logical `AND` of **literals**. A **literals** is either x_i or $\bar{x}_i$. Additionally, no term contains both a variables and its negation. The number of literals of a term is called its **width**.

The Mansour's conjecture states that a polynomial sized DNF may have its Fourier spectrum concentrated on a small collection.

Conjecture 9.2 (Mansour's conjecture)**.** *For $f : \{-1,1\}^n \to \{-1,1\}$ and $\epsilon \in (0, 1/2]$,*

- *If f is computed by a DNF formula of size $s > 1$, then the Fourier spectrum of f is ϵ-concentrated on some family $\mathcal{F}$ of subsets $S \subseteq [n]$ such that $|\mathcal{F}| \leq s^{O(\log(1/\epsilon))}$.*
- *If f is computed by a polynomial sized DNF formula, then the Fourier spectrum of f is ϵ-concentrated on some polynomial sized family $\mathcal{F}$ of subsets $S \subseteq [n]$.*

In this section, we aim to show the following theorem, a weaker result of a DNF with some width $w > 2$.

Theorem 9.3 (Main Theorem)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be computed by a DNF of width $w > 2$. Then For $\epsilon \in (0, 1/2]$, the Fourier spectrum of f is ϵ-concentrated on some family $\mathcal{F}$ of subsets $S \subseteq [n]$ with*

$$|\mathcal{F}| \leq w^{O(w \log(1/\epsilon))}. \tag{9.0.1}$$

If the size of the DNF, $s = poly(n)$, we have $|\mathcal{F}| \leq n^{O(\log \log n)}$.

To prove the above theorem, we need two main tools in the following, whose proofs we will show later.

Theorem 9.4. *For $f : \{-1,1\}^n \to \{-1,1\}$ computed by a DNF of width w and some $\epsilon \in (0, 1/2]$, f is ϵ-concentrated on degree up to $O(w \log(1/\epsilon))$.*

Theorem 9.5. *If $f : \{-1,1\}^n \to \{-1,1\}$ is computed by a DNF of width w, then for all k, we have*

$$\sum_{|U| \leq k} |\hat{f}(U)| \leq 2 \cdot (20w)^k. \tag{9.0.2}$$

9.1 Hastad's Switching Lemma ❖

Definition 9.6 (Random Subset)**.** For $\delta \in [0, 1]$, we say that $\mathcal{J}$ is a **δ-random subset** of $[n]$ if for each element $i \in [n]$, $i \in \mathcal{J}$ with probability δ.

Definition 9.7 (Random Restriction)**.** We define a **δ-random restriction** on $\{-1,1\}^n$ as a pair $(\mathcal{J}|\mathsf{z})$ by first choosing the δ-random subset $\mathcal{J}$ and then choosing $\mathsf{z} \in \{-1,1\}^{\overline{\mathcal{J}}}$ uniformly at random. That is, each coordinate i is **free** if $i \in \mathcal{J}$ and is **fixed** if $i \notin \mathcal{J}$.

Equivalently, each coordinate i is free with probability δ and is fixed to either -1 or $+1$ with probability $(1-\delta)/2$, respectively. Recall the definition of decision tree complexity 7.3, $D(f)$, the minimum depth of any boolean decision tree computing f. We state Hastad's Switching Lemma without proof.

Lemma 9.8 (Hastad's Switching Lemma)**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be computed by a DNF (or CNF) of width at most w and $(\mathcal{J}|\mathsf{z})$ be a δ-random restriction on $\{-1,1\}^n$. Then, for any nonnegative integer k, we have*

$$\mathbb{P}[D(f_{\mathcal{J}|\mathsf{z}}) \geq k] \leq (5\delta w)^k \tag{9.1.1}$$

In the case that $k = 1$, we have that $\mathbb{P}[D(f_{\mathcal{J}|\mathsf{z}})$ is not a constant $] \leq 5\delta w$.

Lemma 9.9. *Let $(\mathcal{J}|\mathsf{z})$ be a δ-random restriction on $\{-1,1\}^n$. For $f : \{-1,1\}^n \to \{-1,1\}$ and some nonnegative integer k, let $\gamma = \mathbb{P}[D(f_{\mathcal{J}|\mathsf{z}}) \geq k]$. We have that f is 3γ-concentrated on degree up to $3k/\delta$.*

Proof. Recall that $\deg(f) \leq D(f)$. Hence, we have

$$\mathbb{P}[\sum_{|S|>k} \hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2 \neq 0] \leq \gamma \tag{9.1.2}$$

By Parseval's Theorem 3.3, we have $\sum_{|S|>k} \hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2 \leq 1$, which implies

$$\mathbb{E}_{\mathcal{J}|\mathsf{z}}[\sum_{|S|>k} \hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] \leq \mathbb{P}[\sum_{|S|>k} \hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2 \neq 0] \leq \gamma \tag{9.1.3}$$

For fixed S, we can write $F_S(\mathsf{z}) = \hat{f}_{*\mathsf{z}}(s)$ as a function of z where $*$ denotes the free bits. As we have seen in the proof of claim 5.10, $\hat{F}_S(\mathsf{z}) = \hat{f}(S \cup T)$ and

$$\mathbb{E}_{\mathsf{z}}[\hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] = \sum_{T \subseteq \overline{\mathcal{J}}} \hat{f}(S \cup T)^2 \tag{9.1.4}$$

Hence, for fixed $\mathcal{J}$, we have that for all $S \subseteq \mathcal{J}$,

$$\mathbb{E}_{\mathsf{z} \in \{-1,1\}^{\overline{\mathcal{J}}}}[\hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] = \sum_{T \subseteq \overline{\mathcal{J}}} \hat{f}(S \cup T)^2 = \sum_{U \subseteq [n]} \hat{f}(U)^2 \mathbf{1}_{U \cap \mathcal{J} = S} \tag{9.1.5}$$

Notice that $f_{\mathcal{J}|\mathsf{z}}$ only depends on the bits in $\mathcal{J}$, which implies that for all $S \subseteq [n], \mathbb{E}_{\mathsf{z}}[\hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \mathbf{1}_{U \cap \mathcal{J} = S}$ and thus

$$\mathbb{E}_{\mathcal{J}|\mathsf{z}}[\hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \mathbb{P}[U \cap \mathcal{J} = S] \tag{9.1.6}$$

Using linearity of expectation,

$$\mathbb{E}_{\mathcal{J}|\mathsf{z}}[\sum_{|S|>k} \hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] = \sum_{|S|>k} \mathbb{E}_{\mathcal{J}|\mathsf{z}}[\hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] = \sum_{|S|>k} \sum_{U \subseteq [n]} \hat{f}(U)^2 \mathbb{P}[U \cap \mathcal{J} = S] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \mathbb{P}[|U \cap \mathcal{J}| > k] \tag{9.1.7}$$

Using Chernoff's Bound, we have that $\mathbb{P}[|U \cap \mathcal{J}| > k] > 1/3$ when $|U| \geq 3k/\delta$. Hence,

$$\mathbb{E}_{\mathcal{J}|\mathsf{z}}[\sum_{|S|>k} \hat{f}_{\mathcal{J}|\mathsf{z}}(S)^2] = \sum_{U \subseteq [n]} \hat{f}(U)^2 \mathbb{P}[|U \cap \mathcal{J}| > k] \geq \sum_{|U| \geq 3k/\delta} \frac{1}{3} \hat{f}(U)^2 \tag{9.1.8}$$

That is, $\sum_{|U| \geq 3k/\delta} \hat{f}(U)^2 \leq 3\delta$. □

proof of Theorem 9.4. Let $\delta = 1/10w, k = C\log(1/\epsilon)$ for some constant C. By Hastad's Switching Lemma 9.8, we have

$$\gamma = \mathbb{P}[D(f_{\mathcal{J}|\mathsf{z}}) \geq k] \leq (5\delta w)^k \leq (\frac{1}{2})^k = \epsilon^C \tag{9.1.9}$$

For large enough C, we have $3\gamma < \epsilon$. Hence, by lemma 9.9, we have f is ϵ-concentrated on degree up to $3k/\delta = 10w3C\log(1/\epsilon) = O(w\log(1/\epsilon))$. □

We state the following lemma without proof.

Lemma 9.10. *Let $f : \{-1,1\}^n \to \{-1,1\}$ and $(\mathcal{J}|\mathsf{z})$ be a δ-random restriction on $\{-1,1\}^n$. Then*

$$\sum_{U \subseteq [n]} \delta^{|U|} \cdot |\hat{f}(U)| \leq \mathbb{E}_{\mathcal{J}|\mathsf{z}}[2^{D(f_{\mathcal{J}|\mathsf{z}})}] \tag{9.1.10}$$

proof of Theorem 9.5. Let $\delta = 1/20w$. By Hastad's Switching Lemma 9.8, we have

$$\mathbb{E}_{\mathcal{J}|\mathsf{z}}[2^{D(f_{\mathcal{J}|\mathsf{z}})}] \leq \sum_{d=0}^{\infty} (\frac{5}{20})^d \cdot 2^d = 2 \tag{9.1.11}$$

Hence, from lemma 9.10, we get

$$2 \geq \mathbb{E}_{\mathcal{J}|\mathsf{z}}[2^{D(f_{\mathcal{J}|\mathsf{z}})}] \geq \sum_{U \subseteq [n]} (\frac{1}{20w})^{|U|} \cdot |\hat{f}(U)| \geq (\frac{1}{20w})^k \sum_{U \subseteq [n]} \cdot |\hat{f}(U)| \tag{9.1.12}$$

□

9.2 Proof of the Main Theorem ❖

Proposition 9.11. *Suppose $f : \{-1,1\}^n \to \mathbb{R}$ is ϵ-concentrated on $\mathcal{F}$ and $g : \{-1,1\}^n \to \mathbb{R}$ satisfies $||f-g||_2^2 \le \delta$. Then, g is $2(\epsilon+\delta)$-concentrated on $\mathcal{F}$.*

Proof. Observe that

$$\sum_{S\notin\mathcal{F}} \hat{g}(S)^2 = \sum_{S\notin\mathcal{F}} [\hat{g}(S) - \hat{f}(S) + \hat{f}(S)]^2 \tag{9.2.1}$$

Using Cauchy-Schwarz inequality, we have

$$[\hat{g}(S) - \hat{f}(S) + \hat{f}(S)]^2 \le 2[\hat{g}(S) - \hat{f}(S)]^2 + \hat{f}(S)^2 \tag{9.2.2}$$

Thus,

$$\begin{aligned}\sum_{S\notin\mathcal{F}} \hat{g}(S)^2 &\le 2\Big(\sum_{S\notin\mathcal{F}} \hat{f}(S)^2 + \sum_{S\notin\mathcal{F}} [\hat{g}(S) - \hat{f}(S)]^2\Big) \le 2\Big(\sum_{S\notin\mathcal{F}} \hat{f}(S)^2 + \sum_{S\subseteq[n]} [\hat{g}(S) - \hat{f}(S)]^2\Big) \\ &= 2\Big(\sum_{S\notin\mathcal{F}} \hat{f}(S)^2 + ||f-g||^2\Big) \le 2(\epsilon+\delta)\end{aligned} \tag{9.2.3}$$

□

Proposition 9.12. *For $f : \{-1,1\}^n \to \{-1,1\}, g : \{-1,1\}^n \to \{-1,1\}$ and some $\epsilon \in (0,1/2]$, f and g are ϵ-close if and only if $||f-g||_2^2 \le 4\epsilon$.*

Proof. $||f-g||^2 = \langle f-g, f-g\rangle = \mathbb{E}_{x\in\{-1,1\}^n}[f(x)-g(x)]^2 = 4\mathbb{P}_{x\in\{-1,1\}^n}[f(x)\neq g(x)] \le 4\epsilon$. □

Lemma 9.13. *For $f : \{-1,1\}^n \to \{-1,1\}$ computed by a DNF of size s and some $\epsilon \in (0,1/2]$, we have f is ϵ-close to a width $\log(s/\epsilon)$ DNF.*

Proof. Let g be the DNF obtained from f by deleting the terms with more than $\log(s/\epsilon)$ literals. Notice that for any fixed term with k literals, the fraction of inputs that satisfies this term is at most 2^{-k}. That is, for $k > \log(s/\epsilon)$, the fraction of inputs that satisfies this term is less than ϵ/s. Therefore,

$$\mathbb{P}_{x\in\{-1,1\}}[f(x)\neq g(x)] \le s\cdot\frac{\epsilon}{s} = \epsilon. \tag{9.2.4}$$

□

proof of theorem 9.3. Let $k = C\cdot w\cdot\log(2/\epsilon)$ for some constant C and $g = \sum_{|S|\le k}\hat{f}(S)\chi_S$. By theorem 9.4, f is ϵ-concentrated on degree up to $O(w\log(1/\epsilon))$. Hence, $\sum_{|S|>k}\hat{f}(S)^2 = \langle f-g\rangle \le \epsilon/2$ for large enough C.
By theorem 9.5, we have $\hat{||}g\hat{||}_1 \le w^{O(w\log(1/\epsilon))}$. By lemma 9.13, we have that g is $\epsilon/2$-concentrated on $\mathcal{F}$ with

$$|\mathcal{F}| \le \Big(\frac{2\hat{||}g\hat{||}_1}{\epsilon}\Big) \le w^{O(w\log(1/\epsilon))} \tag{9.2.5}$$

It remains to show that f is ϵ-concentrated on the same $\mathcal{F}$. Notice that

$$\sum_{S\notin\mathcal{F}}\hat{f}(S)^2 = \sum_{S\notin\mathcal{F},|S|>k}\hat{f}(S)^2 + \sum_{S\notin\mathcal{F},|S|\le k}\hat{f}(S)^2 = \sum_{S\notin\mathcal{F},|S|>k}\hat{f}(S)^2 + \sum_{S\notin\mathcal{F},|S|\le k}\hat{g}(S)^2 \le \epsilon/2+\epsilon/2 = \epsilon \tag{9.2.6}$$

□

Corollary 9.14. *For $f : \{-1,1\}^n \to \{-1,1\}$ computed by a DNF of size s and some $\epsilon \in (0,1/2]$, we have f is $O(\epsilon)$-concentrated on*

$$\mathcal{F} = \{T\subseteq[n] || T| \le \log(\frac{s}{\epsilon})\} \tag{9.2.7}$$

Proof. By lemma 9.13, we have that f is ϵ-close to some width $\log(s/\epsilon)$ DNF g. By proposition 9.12, we have $||f-g||_2^2 \leq 4\epsilon$. Now, using theorem 9.4, we have that g is δ-concentrated on degree up to $O(\log(s/\epsilon)\log(1/\delta))$. By proposition 9.11, we have that f is $2(\delta+4\epsilon)$-concentrated on $\mathcal{F}$. □

We can improve this corollary if we replace theorem 9.4 by theorem 9.3 in the proof we have shown, which gives us the following.

Corollary 9.15. *For $f:\{-1,1\}^n \to \{-1,1\}$ computed by a DNF of size s and some $\epsilon \in (0,1/2]$, we have f is $O(\epsilon)$-concentrated on $\mathcal{F}$ with*

$$|\mathcal{F}| \leq (\log(s/\epsilon))^{O(\log(s/\epsilon)\log(1/\epsilon))}. \tag{9.2.8}$$

Remark 9.16. For constant ϵ, we have

$$|\mathcal{F}| \leq (\log(s))^{O(\log(s))} = s^{O(\log\log(s))}. \tag{9.2.9}$$

If we further have that $s = \text{poly}(n)$, then

$$|\mathcal{F}| \leq n^{O(\log\log(n))}. \tag{9.2.10}$$

10 Influence and Sensitivity

For the first half of this section, we restrict our attention to Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$. For the second half, we will move on to more general definitions for real valued functions $f : \{-1,1\}^n \to \mathbb{R}$.

Notation 10.1. We use x^i to denote the string x whose ith bit is flipped.

Definition 10.2. For a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we say that f is **sensitive** to the ith bit on input x if $f(x) \neq f(x^i)$.

Definition 10.3. The **sensitivity** of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ **on input** $x \in \{-1,1\}^n$ is defined as

$$s(f,x) := \#\text{ of coordinates } i \text{ such that } f(x) \neq f(x^i) \tag{10.0.1}$$

The **sensitivity** of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$s(f) := \max_{x \sim \{-1,1\}^n} s(f,x) \tag{10.0.2}$$

The **average sensitivity** of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$\text{Avg Sensitivity}(f) := \mathbb{E}_{x \sim \{-1,1\}^n}[s(f,x)] = 2^{-n} \sum_{x \sim \{-1,1\}^n} s(f,x) \tag{10.0.3}$$

Example 10.4. $s(\mathtt{OR}_n, 0...0) = n, \quad s(\mathtt{OR}_n, x) = 1$ where $|x| = 1, \quad s(\mathtt{OR}_n, x) = 0$ where $|x| > 1$.

Example 10.5. $s(\mathtt{XOR}_n, x) = n$.

Definition 10.6. The **influence** of the i-th bit on $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$\mathbf{Inf}_i[f] = \mathbb{P}_{x \sim \{-1,1\}^n}[f(x) \neq f(x^i)] = \frac{|\{x : f(x) \neq f(x^i)\}|}{2^n} \tag{10.0.4}$$

The **total influence** of $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$\mathbf{I}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f] \tag{10.0.5}$$

Example 10.7. For $\mathtt{OR}_n$, we have for any i,

$$\mathbf{Inf}_i[\mathtt{OR}_n] = 2^{-n+1} \tag{10.0.6}$$

Similarly, for $\mathtt{AND}_n$, we have for any i,

$$\mathbf{Inf}_i[\mathtt{AND}_n] = 2^{-n+1} \tag{10.0.7}$$

Thus,

$$\mathbf{I}[\mathtt{OR}_n] = \mathbf{I}[\mathtt{AND}_n] = n2^{-n+1} \tag{10.0.8}$$

Example 10.8. As for the parity function $\mathtt{XOR}_n = \chi_{[n]}$, we have for any i,

$$\mathbf{Inf}_i[\chi_{[n]}] = 1 \tag{10.0.9}$$

and thus,

$$\mathbf{I}[\chi_{[n]}] = n \tag{10.0.10}$$

Example 10.9. We can use a Boolean cube to help visualise the influence of $\mathtt{Maj}_3$, as shown in Figure 10.1. We let each Boolean string represent the input and draw an edge between two strings if one string x can be obtained by flipping one bit from the other string y. Hence, if $f(x) \neq f(y)$, we mark this edge.
Hence, we obtain that for all $i, \mathbf{Inf}_i[\mathtt{Maj}_3] = 1/2$ and $\mathbf{I}[\mathtt{Maj}_3] = 3/2$.

Remark 10.10. From the viewpoint of Boolean cubes, we have that $\mathbf{Inf}_i[f]$ equals the fraction of marked edges in the ith dimension, or equivalently

$$\mathbf{Inf}_i[f] = \frac{\#\text{ of pairs connected by marked edges}}{2^{n-1}} \tag{10.0.11}$$

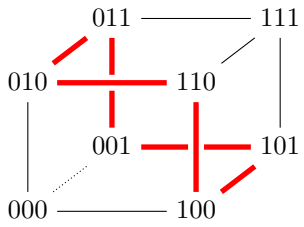


Figure 10.1: Boolean Cube for $\texttt{Maj}_3$

10.1 Influence and Derivatives ❖

We now adapt a more general definition of influence.

Definition 10.11. The **influence** of the i-th bit on $f : \{-1,1\}^n \to \mathbb{R}$ is defined as

$$\mathbf{Inf}_i[f] = \sum_{S \text{ s.t. } i \in S} \hat{f}(S)^2 \tag{10.1.1}$$

The **total influence** of $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$\mathbf{I}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f] \tag{10.1.2}$$

Definition 10.12. For $f : \{-1,1\}^n \to \mathbb{R}$, we define the **derivative in the i-th direction** as

$$D_{i,f}(x) = \frac{f(x^{i\to 1}) - f(x^{i\to -1})}{2} \tag{10.1.3}$$

For Boolean valued $f : \{-1,1\}^n \to \{-1,1\}$, we have

$$D_{i,f}(x) = \begin{cases} \pm 1 & \text{if } f(x) \neq f(x^i) \\ 0 & \text{if } f(x) = f(x^i) \end{cases} \tag{10.1.4}$$

Note that $D_{i,f}(x)$ has only $n-1$ variables.

Lemma 10.13. *For a set $T \subseteq [n] \setminus \{i\}$ for some coordinate i, we have*

$$\hat{D}_{i,f}(T) = \hat{f}(T \cup \{i\}) \tag{10.1.5}$$

Proof. Suppose the coordinate i is given. Let $S \subseteq [n]$ be a set that contains i. Then,

$$\begin{aligned} \hat{f}(S) &= \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x)\chi_S(x) \\ &= \frac{1}{2^n} \sum_{x \in \{-1,1\}^n, x_i = 1} f(x)\chi_S(x) + f(x^i)\chi_S(x^i) \\ &= \frac{1}{2^n} \sum_{x \in \{-1,1\}^n, x_i = 1, f(x) \neq f(x')} \chi_{S\setminus\{i\}}(f(x) - f(x^i)) \\ &= \frac{1}{2^{n-1}} \sum_{y \in \{-1,1\}^{n-1}} \chi_{S\setminus\{i\}} D_{i,f}(y) \\ &= \hat{D}_{i,f}(S \setminus \{i\}) \end{aligned} \tag{10.1.6}$$

□

Notation 10.14. We use the indicator function

$$\mathbf{1}_{f(x)\neq f(x^i)}(x) = \begin{cases} 1 & \text{if } f(x) \neq f(x^i) \\ 0 & \text{if } f(x) = f(x^i) \end{cases} \tag{10.1.7}$$

Notice that $D_{i,f}(x)^2 = \mathbf{1}_{f(x)\neq f(x^i)}(x)$.

Proposition 10.15. *For $f : \{-1,1\}^n \to \{-1,1\}$, the two definitions of influence on the i-th bit is equivalent. In other words,*

$$\mathbf{Inf}_i[f] = \mathbb{P}_{x\sim\{-1,1\}^n}[f(x) \neq f(x^i)] = \sum_{S\subseteq[n] \ s.t. \ i\in S} \hat{f}(S)^2 \tag{10.1.8}$$

Proof.

$$\begin{aligned} \mathbf{Inf}_i[f] &= \mathbb{P}_{x\sim\{-1,1\}^n}[f(x) \neq f(x^i)] \\ &= \mathbb{E}_{x\sim\{-1,1\}^n}[\mathbf{1}_{f(x)\neq f(x^i)}(x)] \\ &= \mathbb{E}_{x\sim\{-1,1\}^n}[D_{i,f}(x)^2] \\ &= \frac{1}{2^{n-1}} \sum_{x\sim\{-1,1\}^n, x_i=1} \frac{D_{i,f}(x)^2 + D_{i,f}(x^i)^2}{2} \end{aligned} \tag{10.1.9}$$

Notice that $D_{i,f}(x)^2 = D_{i,f}(x^i)^2$, which gives us

$$\begin{aligned} \mathbf{Inf}_i[f] &= \frac{1}{2^{n-1}} \sum_{x\sim\{-1,1\}^n, x_i=1} \frac{D_{i,f}(x)^2 + D_{i,f}(x^i)^2}{2} \\ &= \mathbb{E}_{y\sim\{-1,1\}^{n-1}}[D_{i,f}(y)^2] \\ &= \langle D_{i,f}, D_{i,f} \rangle \\ &= \sum_{T\subseteq[n]\setminus\{i\}} \hat{D}_{i,f}(T)^2 \end{aligned} \tag{10.1.10}$$

that does not depend on the i-th bit. Hence, by lemma 10.13, we obtain

$$\mathbf{Inf}_i[f] = \sum_{T\subseteq[n]\setminus\{i\}} \hat{D}_{i,f}(T)^2 = \sum_{S\subseteq[n] \text{ s.t. } i\in S} \hat{f}(S)^2 \tag{10.1.11}$$

□

Proposition 10.16. *For any Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, we have*

$$\mathbf{I}[f] = \textit{Avg Sensitivity}(f) \tag{10.1.12}$$

Proof.

$$\begin{aligned} \mathbf{I}[f] &= \sum_{i=1}^n \mathbb{P}_{x\sim\{-1,1\}^n}[f(x) \neq f(x^i)] \\ &= \sum_{i=1}^n \mathbb{E}_{x\sim\{-1,1\}^n}[\mathbf{1}_{f(x)\neq f(x^i)}(x)] \\ &= \mathbb{E}_{x\sim\{-1,1\}^n}[\sum_{i=1}^n \mathbf{1}_{f(x)\neq f(x^i)}(x)] \\ &= \mathbb{E}_{x\sim\{-1,1\}^n}[s(f,x)] = \text{Avg Sensitivity}(f) \end{aligned} \tag{10.1.13}$$

□

Theorem 10.17. *For $f : \{-1,1\}^n \to \mathbb{R}$, $\mathbf{I}[f] = \sum_{S\in[n]} |S| \cdot \hat{f}(S)^2$.*

Proof.

$$\mathbf{I}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f] = \sum_{i=1}^{n} \sum_{S \text{ s.t. } i\in S} \hat{f}(S)^2 = \sum_{S\in[n]} |S| \cdot \hat{f}(S)^2 \tag{10.1.14}$$

□

Corollary 10.18. *$f : \{-1,1\}^n \to \{-1,1\}$, $\mathbf{I}[f] \leq \deg(f)$.*

Proof.

$$\mathbf{I}[f] \sum_{S\in[n]} |S| \cdot \hat{f}(S)^2 \leq \max_{S\subseteq[n] \text{ s.t. } \hat{f}(S)\neq 0} |S| \sum_{S\in[n]} \cdot \hat{f}(S)^2 \leq \deg(f) \tag{10.1.15}$$

□

Corollary 10.19. *For $f : \{-1,1\}^n \to \mathbb{R}$, f is ϵ-concentrated up to degree $\mathbf{I}[f]/\epsilon$.*

Proof. Note that

$$\mathbf{I}[f] \geq \sum_{S\subseteq[n] \text{ s.t. } |S|>\mathbf{I}[f]/\epsilon} |S| \cdot \hat{f}(S)^2 \geq \frac{\mathbf{I}[f]}{\epsilon} \sum_{S\in[n] \text{ s.t. } |S|>\mathbf{I}[f]/\epsilon} \hat{f}(S)^2 \tag{10.1.16}$$

which gives us

$$\sum_{S\in[n] \text{ s.t. } |S|>\mathbf{I}[f]/\epsilon} \hat{f}(S)^2 \leq \epsilon \tag{10.1.17}$$

□

10.2 Mean and Variance ❖

For $f : \{-1,1\}^n \to \mathbb{R}$, think of $f(x)$ as a real valued random variable. The **mean** of f is given by

$$\mathbb{E}[f] = \mathbb{E}_{x\in\{-1,1\}^n}[f(x)] \tag{10.2.1}$$

The **variance** of f is given by

$$Var[f] = \langle f - \mathbb{E}[f], f - \mathbb{E}[f]\rangle = \mathbb{E}[f^2] - \mathbb{E}[f]^2 \tag{10.2.2}$$

Lemma 10.20. *For $f : \{-1,1\}^n \to \mathbb{R}$, we have $\mathbb{E}[f] = \hat{f}(\emptyset), Var[f] = \sum_{S\subseteq[n],S\neq\emptyset} \hat{f}(S)^2$.*

Proof.

$$\mathbb{E}[f] = \langle f, 1\rangle = \langle f, \chi_\emptyset\rangle = \hat{f}(\emptyset) \tag{10.2.3}$$

$$Var[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{S\subseteq[n]} \hat{f}(S)^2 - \hat{f}(\emptyset)^2 = \sum_{S\subseteq[n],S\neq\emptyset} \hat{f}(S)^2 \tag{10.2.4}$$

□

Lemma 10.21. *For $f : \{-1,1\}^n \to \{-1,1\}$, let $p = \mathbb{P}_{x\in\{-1,1\}^n}[f(x) = -1]$. We have $\mathbb{E}[f] = 1 - 2p, Var[f] = 4p(1-p)$.*

Proof.

$$\mathbb{E}[f] = 1 \cdot \mathbb{P}_{x\in\{-1,1\}^n}[f(x) = 1] + (-1) \cdot \mathbb{P}_{x\in\{-1,1\}^n}[f(x) = -1] = (1-p) - p = 1 - 2p \tag{10.2.5}$$

$$Var[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = 1 - (1-2p)^2 = 4p(1-p) \tag{10.2.6}$$

□

Theorem 10.22 (Poincare Inequality)**.** *For $f : \{-1,1\}^n \to \mathbb{R}$, we have $Var[f] \leq \mathbf{I}[f]$.*

Proof. Because $|S| \geq 1$ if $S \neq \emptyset$, we have

$$Var[f] = \sum_{S \subseteq [n], S \neq \emptyset} \hat{f}(S)^2 \leq \sum_{S \subseteq [n]} |S| f(S)^2 = \mathbf{I}[f] \tag{10.2.7}$$

□

Corollary 10.23. *For $f : \{-1,1\}^n \to \mathbb{R}$, we have*

$$\frac{1}{n} Var[f] \leq \max_i \mathbf{Inf}_i[f] \leq Var[f] \tag{10.2.8}$$

Proof. The first inequality comes from the Poincare Inequality (see above) as follows:

$$\frac{1}{n} Var[f] \leq \frac{1}{n} \mathbf{I}[f] \leq \frac{1}{n} \sum_i \mathbf{Inf}_i[f] \leq \max_i \mathbf{Inf}_i[f] \tag{10.2.9}$$

To see the second inequality, we observe that for all i, by proposition 10.15 and lemma 10.20,

$$\mathbf{Inf}_i[f] = \sum_{S \subseteq [n] \text{ s.t. } i \in S} \hat{f}(S)^2 \leq \sum_{S \subseteq [n], S \neq \emptyset} \hat{f}(S)^2 = Var[f] \tag{10.2.10}$$

□

10.3 Monotone Functions ❖

Definition 10.24. We say that a function $f : \{0,1\}^n \to \{0,1\}$ is **monotone** if for $x, y \in \{0,1\}^n$ such that $x \leq y$, (i.e., for all $i, x_i \leq y_i$), then $f(x) \leq f(y)$.

Notice that for monotone $f : \{-1,1\}^n \to \{-1,1\}$, we have $D_{i,f}(x) = \mathbf{1}_{f(x) \neq f(x^i)}(x)$.

Example 10.25. $\mathtt{AND}_n, \mathtt{OR}_n, \mathtt{Maj}_n$ are monotone. $\mathtt{XOR}_n$ is not monotone.

Proposition 10.26. *For monotone $f : \{-1,1\}^n \to \{-1,1\}$,*

$$\mathbf{Inf}_i[f] = \hat{f}(\{i\}) \tag{10.3.1}$$

Proof. By lemma 10.20 and lemma 10.13, we have

$$\mathbf{Inf}_i[f] = \mathbb{E}[\mathbf{1}_{f(x) \neq f(x^i)}(x)] = \mathbb{E}[D_{i,f}(x)] = \hat{D}_{i,f}(\emptyset) = \hat{f}(\{i\}) \tag{10.3.2}$$

□

11 Voting Rules and Coalitions

A Boolean function $f : \{0,1\}^n \to \{0,1\}$ can be thought of as a **voting rule** or **social choice function** for an election with 2 candidates and n voters. The output of f can be thought of as the winner. The goal of a good voting rule is a balanced function where each bit has not too large influence. We say that a function is balanced if $\mathbb{P}_{x\sim\{0,1\}^n}[f(x) = 1] = 1/2$.

Example 11.1. As we have seen in example 10.7, `AND` and `OR` has equally small individual influence and these functions are far from being balanced.

Example 11.2. The **dictator function**, i.e., $f(x) = x_i$ for some coordinate i is balanced; however, for a dictator function $f(x) = x_i$

$$\mathbf{Inf}_i(x_i) = 1 \quad \mathbf{Inf}_j(x_i) = 0 \forall i \neq j \tag{11.0.1}$$

This means that only the ith bit is influential.

Example 11.3. We say a function $f : \{0,1\}^n \to \{0,1\}$ is a **tribes** function of width w and size s where $n = ws$ if

$$\mathtt{OR}_s(\mathtt{AND}_w(x^{(1)}), ..., \mathtt{AND}(x^{(s)})) \tag{11.0.2}$$

where $x^{(i)} \in \{0,1\}^w$. Notice that

$$\mathbb{P}_x[Tribes_{w,s}(x) = 0] = (1 - \frac{1}{2^w})^s \tag{11.0.3}$$

If we set $s := \lceil 2^w \ln 2 \rceil$, then we have that $\mathbb{P}_{x\in\{0,1\}^n}[Tribes_{w,s}(x) = 0] \approx 1/2$. Hence, Tribes function is close to balanced. Moreover, the individual influence of each bit is also not too large. Specifically, consider a bit k in the ith tribe. Notice that with probability $1/2^{w-1}$ all other bits in the ith tribe votes `TRUE` and with probability $(1 - 1/2^{w-1})^{s-1}$ all other tribes votes `FALSE`. Then

$$\begin{aligned}
\mathbf{Inf}_k[Tribes_{w,s}] &= \mathbb{P}_x(Tribes_{w,s}(x) \neq Tribes_{w,s}(x^k)) \\
&= \frac{1}{2^{w-1}} \cdot (1 - \frac{1}{2^{w-1}})^{s-1} \\
&= \frac{1}{2^{w-1}} \cdot (1 - \frac{1}{2^{w-1}})^{-1} \cdot \mathbb{P}_x[Tribes_{w,s}(x) = 0] \\
&= \frac{2}{2^w - 1} \cdot \mathbb{P}_x[Tribes_{w,s}(x) = 0] \\
&\approx \frac{2}{2^w - 1} \cdot \frac{1}{2} \approx \frac{1}{2^w - 1}
\end{aligned} \tag{11.0.4}$$

Notice that we set $s := \lceil 2^w \ln 2 \rceil$ and $ws = n$, then

$$\mathbf{Inf}_k[Tribes_{w,s}] \approx \frac{1}{2^w - 1} \approx \frac{w \ln 2}{n} \approx \frac{\ln n}{n} \tag{11.0.5}$$

Thus, each individual bit also has not too large influence.

11.1 Symmetric Functions and Transitive Functions ❖

Definition 11.4. We say that a function $f : \{-1,1\}^n \to \{-1,1\}$ is **symmetric** if $f(x^\pi) = f(x)$ for all permutation π. In other words, $f(x)$ only depends on the number of 1s in x and is independent of the way the bits are permuted.

Example 11.5. `AND, OR, PARITY, MAJORITY` are all symmetric.

Transitive function are a generalisation of symmetric functions.

Definition 11.6. We say that a function $f : \{-1,1\}^n \to \{-1,1\}$ is **transitive** if for all $i, j \in [n]$ there exists a permutation π taking i to j such that $f(x^\pi) = f(x)$.

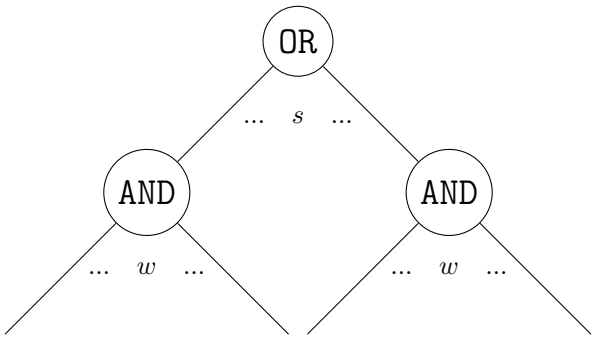


Figure 11.1: Tribes Function

Intuitively, a function f is transitive if for any two $i, j \in [n]$ are "equivalent".

Theorem 11.7. *For a transitive function $f : \{-1,1\}^n \to \{-1,1\}$,*

$$\mathbf{Inf}_i[f] = \mathbf{Inf}_j[f] \quad \forall i, j \in [n] \tag{11.1.1}$$

Example 11.8. `AND, OR, PARITY, MAJORITY` are all symmetric, which implies that they are transitive.

Example 11.9. Tribes function is transitive but not symmetric.

Example 11.10. Dictator function is neither transitive nor symmetric.

11.2 Bounds of Maximum Individual Influence ❖

Recall that as a corollary of Poincare Inequality, we obtained the following bound 10.23.

$$\frac{1}{n} Var[f] \leq \max_i \mathbf{Inf}_i[f] \leq Var[f]$$

As for functions that are "somewhat" balanced, we have obtained asymptotically tight bounds on maximum individual influence.

Theorem 11.11 ([KKL88])**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$. Then there exists a coordinate $i \in [n]$ such that*

$$\mathbf{Inf}_i[f] \geq \Omega(\frac{\ln n}{n}) \cdot Var[f] \tag{11.2.1}$$

In other words, if f is not "too unbalanced", for example, neither p nor $1-p$ is too small, then we have

$$\mathbf{Inf}_i[f] \geq \Omega(\frac{\ln n}{n})$$

Notice that this is tight and the Tribes function serves as an example.

A generalisation of this result is given by Talagrand.

Theorem 11.12 ([Tal94])**.**

$$\sum_{i=1}^{n} \frac{\mathbf{Inf}_i[f]}{\log\left(\frac{1}{\mathbf{Inf}_i[f]}\right)} \geq \Omega(Var[f]) \tag{11.2.2}$$

Notice that this theorem implies the Kahn-Kalai-Linial theorem 11.11. Specifically,

$$\max_{i\in[n]} \frac{\mathbf{Inf}_i[f]}{\log\left(\frac{1}{\mathbf{Inf}_i[f]}\right)} \geq \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{Inf}_i[f]}{\log\left(\frac{1}{\mathbf{Inf}_i[f]}\right)} \geq \frac{1}{n}\Omega(Var[f])$$

Another bound relates the decision tree.

Theorem 11.13 ([OSSS05]). *If $f : \{-1,1\}^n \to \{-1,1\}$ is "almost balanced" and has a decision tree with depth d, then*

$$\max_{i\in[n]} \mathbf{Inf}_i[f] \geq \Omega(\frac{1}{d}) \tag{11.2.3}$$

11.3 Influence of Coalitions ❖

Let $S \subseteq [n]$, we denote the influence of this coalition

$$\mathbf{Inf}_S[f] = \mathbb{P}_{y\in\{-1,1\}^{n-|S|}}[f_{s,y} \text{ is not a constant}] \tag{11.3.1}$$

where $f_{s,y}$ is a sub-function over variables in S with y fixed. A corollary of the Kahn-Kalai-Linial theorem 11.11 gives a bound on the concentration of influence

Corollary 11.14. *If $f : \{-1,1\}^n \to \{-1,1\}$ is "almost balanced", then for all $\epsilon > 0$, there exists a coalition $S \subseteq [n]$ of size*

$$|S| \leq O(\log(\frac{1}{\epsilon})\frac{n}{\log n}) \tag{11.3.2}$$

such that $\mathbf{Inf}_S[f] \geq 1 - \epsilon$.

This bound is conjectured to be tight. The following result of Ajtai and Linial shows that the bound is "almost" tight.

Theorem 11.15 ([AL93]). *There exists a balanced function f such that for all $S \subseteq [n]$ with $|S| \leq o(\frac{n}{\log^2 n})$,*

$$\mathbf{Inf}_S[f] \leq o(1) \tag{11.3.3}$$

Such function f is resilient against "large" coalitions that would like to dominate the influence. However, no explicit f is known.

Example 11.16. Even though the tribes function is quite balanced and resilient against influential individuals, it is not very resilient against coalitions. Notice that there exists a set $S, |S| = O(\log n)$ such that

$$\mathbf{Inf}_S[f] \geq \Omega(1) \tag{11.3.4}$$

Example 11.17. The `MAJORITY` function is resilient against coalitions with size at most $n^{1/2-\epsilon}$ for some $\epsilon > 0$. The `ITERATED MAJORITY` function is resilient against coalitions with size at most $n^{\log_3 2}$.

11.4 Juntas ❖

Another class of functions of interest are those with small total influences.

Definition 11.18. A function $f : \{-1,1\}^n \to \{-1,1\}$ is called a **k-junta** for $k \in \mathbb{N}$ of it depends on at most k of its input variables.

Notice that for each of the k variables, its influence is at most 1. Hence, we obtain the following proposition.

Proposition 11.19. *If $f : \{-1,1\}^n \to \{-1,1\}$ is a k-junta, then $\mathbf{I}[f] \leq k$.*

The following theorem states that a function $f : \{-1,1\}^n \to \{-1,1\}$ is ϵ-close to a $2^{O(\mathbf{I}[f]/\epsilon)}$-junta for any $0 < \epsilon < 1$.

Theorem 11.20 (Friedgut's Junta Theorem). *For a function $f : \{-1,1\}^n \to \{-1,1\}$, and any $0 < \epsilon < 1$, there exists a function $g : \{-1,1\}^n \to \{-1,1\}$ such that*

- *g is a $2^{O(\mathbf{I}[f]/\epsilon)}$-junta.*
- *$\mathbb{P}_x[f(x) \neq g(x)] \leq \epsilon$.*

12 Isoperimetric Problem for Graphs

The classical isoperimetric problem is to find, among all closed curves of a given length, the one which encloses the maximum area. The analogue for graphs is natural. Consider a graph $G = (V, E)$ and a subset of vertices $S \subseteq V$. Let F be the set of *all* edges in E that connect the vertices within S. We say that $H = (S, F)$ is the *induced subgraph with vertex set* S. We say the *edge boundary of* H is the set of edges joining a vertex in S to a vertex outside of S. We say the *vertex boundary of* H is the set of vertices outside of S, which are joined to some vertices in S. The isoperimetric problem for graphs is to find a set S whose induced subgraph has the smallest sized boundary.
Consider the n-dimensional Boolean hypercube, in which two vertices are connected by an edge if and only if they differ in exactly 1 coordinate. Harper [Har64] [Har66] has shown that

- Among all a subgraphs with $m = 2^k$ vertices, for some $k \in \mathbb{N}$, the k-dimensional sub-cube has the smallest possible edge boundary.
- Among all a subgraphs with $m = \sum_{i=0}^{r} \binom{n}{i}$ vertices, for some $r \in \mathbb{N}$, the Hamming ball of radius r has the smallest possible vertex boundary.

12.1 Edge Isoperimetric Inequalities ❖

One question of interest is to give lower bounds on the number of edges in the edge boundary for subgraphs with a given number m of vertices.

Notation 12.1. Let $S \subseteq V$ be a subset of vertices. We write $E(S, \bar{S})$ to denote set of edges joining a vertex in S to a vertex outside of S.

The following corollary stresses that the Poincare Inequality is an isoperimetric inequality.

Corollary 12.2. *Let $S \subseteq \{-1, 1\}^n$ have $|S| = \alpha 2^n$. Then $|E(S, \bar{S})| \geq 2\alpha(1 - \alpha)2^n$*

Proof. Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be a function such that $f(x) = -1$ for $x \in S$ and $f(x) = 1$ otherwise.
By remark 10.10, we have $\mathbf{I}[f] = |E(S, \bar{S})|2^{1-n}$. By lemma 10.21 the variance of f is given by $\mathrm{Var}[f] = 4\alpha(1 - \alpha)$. The Poincare Inequality 10.22 gives us $\mathbf{I}[f] \geq \mathrm{Var}[f]$. Hence $|E(S, \bar{S})| \geq 2\alpha(1 - \alpha)2^n$. □

Notice that Poincare Inequality is not very strong. In fact, in [Har64], Harper gives a classical isoperimetric inequality:

$$|E(S, \bar{S})| \geq |S| \log(\frac{2^n}{|S|}) \tag{12.1.1}$$

By replacing $|S|$ with $\alpha 2^n$, we obtain the following theorem.

Theorem 12.3. *For $f : \{-1, 1\}^n \to \{-1, 1\}$ with $\alpha = \mathbb{P}_x[f(x) = -1]$,*

$$\mathbf{I}[f] = \frac{|E(S, \bar{S})|}{2^{n-1}} \geq \frac{|S| \log(2^n/|S|)}{2^{n-1}} = \frac{\alpha 2^n \log(2^n/(\alpha 2^n))}{2^{n-1}} = 2\alpha \log(1/\alpha) \tag{12.1.2}$$

Notice that this inequality is tight, as the equality holds for the $(n - k)$-dimensional sub-cube for $|S| = 2^{n-k}$. That is, $|E(S, \bar{S})| = 2^{n-k}k = |S| \log(2^n/|S|)$.

13 Sensitivity Conjecture

13.1 Complexity Measures ❖

Notation 13.1. For $x \in \{0,1\}^n$ and $i \in [n]$, we use x^i to denote the string x with the ith bit flipped. For $x \in \{0,1\}^n$ and $S \subseteq [n]$, we use x^S to denote the string x with all bits in S flipped. For this section, it is convenient to discuss the concepts in $\{0,1\}$ setting.

Recall the definition of sensitivity,

Definition 13.2. The **sensitivity** of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ **on input** $x \in \{0,1\}^n$ is defined as

$$s(f,x) := \#\text{ of coordinates } i \text{ such that } f(x) \neq f(x^i) \tag{13.1.1}$$

The **sensitivity** of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is defined as

$$s(f) := \max_{x \sim \{0,1\}^n} s(f,x) \tag{13.1.2}$$

We define the **1-sided measures** of sensitivity by

$$s_0(f) := \max_{x \in f^{-1}(0)} s(f,x) \quad s_1(f) := \max_{x \in f^{-1}(1)} s(f,x) \tag{13.1.3}$$

For a boolean function f, let CREW(f) denote the number of steps needed to needed to compute f on a concurrent read, exclusive write parallel random access machine (CREW PRAM) with an unlimited number of processors and memory cells). In 1982, Cook, Dwork and Reischuk found that the CREW PRAM complexity of f could be lower bounded by the logarithm of its sensitivity.

Theorem 13.3 ([CD82], [Rei82])**.** CREW(f) $\geq \Omega(\log s(f))$.

Definition 13.4 (Block Sensitivity)**.** The **block sensitivity** of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ **on input** $x \in \{0,1\}^n$ is defined as the maximum number of mutually disjoint blocks $B_1, B_2, ..., B_t \subseteq [n]$ such that $f(x) \neq f(x^{B_i})$ for all $i \in [t]$. The **block sensitivity** of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is defined as

$$bs(f) := \max_{x \sim \{0,1\}^n} bs(f,x) \tag{13.1.4}$$

Similarly, the 1-sided measures of block sensitivity are

$$bs_0(f) := \max_{x \in f^{-1}(0)} bs(f,x) \quad bs_1(f) := \max_{x \in f^{-1}(1)} bs(f,x) \tag{13.1.5}$$

Recall the definition of decision tree complexity 7.3, $D(f)$, the minimum depth of any boolean decision tree computing f. In 1989, Nisan linked the bound of CREW(f) with decision tree complexity and block sensitivity.

Theorem 13.5 ([Nis89])**.** *For any function f,* CREW(f) $= \Theta(\log D(f)) = \Theta(\log bs(f))$.

Example 13.6. Consider the OR function: $\texttt{OR}(x_1, ..., x_n) = x_1 \vee x_2 \vee ... \vee x_n$. We have

- For the 0-input, when $x = 00...0$, $s(\texttt{OR}, x) = n$.
- For all inputs x with Hamming weight 1, we have $s(\texttt{OR}, x) = 1$.
- For all other inputs x, $s(\texttt{OR}, x) = 0$.

Hence, $s_0(\texttt{OR}) = n, s_1(\texttt{OR}) = 1, s(\texttt{OR}) = \max\{s_0(\texttt{OR}), s_1(\texttt{OR})\} = n$. As for block sensitivity,

- For the 0-input, when $x = 00...0$, $bs(\texttt{OR}, x) = n$.
- For all other inputs x, $bs(\texttt{OR}, x) = 1$.

Hence, $bs_0(\texttt{OR}) = n, bs_1(\texttt{OR}) = 1, bs(\texttt{OR}) = \max\{bs_0(\texttt{OR}), bs_1(\texttt{OR})\} = n$.

Notice that for any function f, $s(f) \leq bs(f)$ as block sensitivity generalises sensitivity. Moreover, for monotone functions, we have $s(f) = bs(f)$.

Theorem 13.7 ([Nis89])**.** *For any monotone function $f, s(f) = bs(f)$.*

What about other functions? The best known separation between sensitivity and block sensitivity of a function remained quadratic since 1995. Various constructions have been given below. Notice that a separation that achieves a constant greater than 1 will imply a *superquadratic* separation.

- [Rub95] $bs(f) = \frac{1}{2}s(f)^2$.
- [Cha05] $bs(f) = \frac{1}{4}s(f)^2$.
- [Vir11] $bs(f) = \frac{1}{2}s(f)^2 + \frac{1}{2}s(f)$.
- [AS11] $bf(s) = \frac{2}{3}s(f)^2 - \frac{1}{3}s(f)$.
- [GSTW16] $bs(f) = \frac{1}{2}s(f)^2 + \frac{1}{2}s(f)$.
- [CG18] $bs(f) \geq \Omega(s(f)^2)$ with constant $1/4, 1/2, 2/3$.

Notation 13.8. For a binary string $x \in \{0,1\}^n$ and a subset $S \subseteq [n]$, we write $x|_S$ be the values of x_i on the coordinates $i \in S$.

Definition 13.9 (Certificate Complexity)**.** A **certificate** of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ **on input** $x \in \{0,1\}^n$ is a subset $S \subseteq [n]$ such that for all $y \in \{0,1\}^n$, if $x|_S = y|_S$, then $f(x) = f(y)$. That is, f is constant if the bits in S are fixed to $x_i, \forall i \in S$.
The **certificate complexity** of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ **on input** $x \in \{0,1\}^n$, denoted by $C(f,x)$, is the smallest size of a certificate of f on x. The **certificate complexity** of f is defined as

$$C(f) := \max_{x \sim \{0,1\}^n} C(f,x) \tag{13.1.6}$$

Similarly,

$$C_0(f) := \max_{x \in f^{-1}(0)} C(f,x) \quad C_1(f) := \max_{x \in f^{-1}(1)} C(f,x) \tag{13.1.7}$$

Example 13.10. For the function $\texttt{OR}_6$ on $x = (101000)$, we have $\{x_3\}$ is a certificate.

Theorem 13.11 ([Nis89])**.** *For any monotone function $f, s(f) = bs(f) = C(f)$.*

Definition 13.12 (Fourier Degree)**.** We say that a polynomial $p : \mathbb{R}^n \to \mathbb{R}$ *represents* a Boolean function $f : \{0,1\}^n \to \{0,1\}$ if for all $x \in \{0,1\}^n$,

$$p(x) = f(x) \tag{13.1.8}$$

The **(Fourier) degree** of f, denoted $\deg(f)$, is the degree of the unique multilinear polynomial that represents f.

Recall that we denote $D(f)$ as the minimum depth of any boolean decision tree computing f. By definition, we see that $s(f) \leq bs(f) \leq C(f) \leq D(f)$. It also turns out that $bs(f), C(f), D(f)$ and $\deg(f)$ are polynomially related. We list some of the results below.

- [NS92] $\deg(f) \leq D(f)$.
- [NS92, Tal13] $bs(f) \leq \deg(f)^2$.
- [BI87, Tar89, AUY83] $D(f) \leq C_0C_1(f) \leq C(f)^2$.
- [Nis89] $C(f) \leq bs(f)^2$.

13.1.1 The Sensitivity Conjecture

In [NS92], Nisan and Szegedy asks the following question.

Question 13.13. Is it true that for every Boolean function $f, bs(s) \leq poly(s(f))$?

Ambainis, Gao, Mao, Sun and Zuo have shown the following.

Theorem 13.14 ([AGM+13]). $bs(f) \leq s(f)2^{s(f)-1}$.

Recently, Huang proved the sensitivity conjecture by combining the Gotsman-Lineal Theorem and results from linear algebra, which we will show later.

Theorem 13.15 ([Hua19]). $\deg(f) \leq s(f)^2$.

Because $bs(f) \leq \deg(f)^2$, we obtain $bs(f) \leq s(f)^4$. Also, Laplante, Naserasr and Sunny have shown the following

Theorem 13.16 ([LNS20]). $\deg(f) \leq s_0(f)s_1(f)$.

13.2 Gotsman-Lineal Theorem ❖

Notation 13.17. For a graph $H = (V, E)$, we let $\Delta(H)$ denote the maximum degree of H.
For an induced subgraph $G = (V_G, E_G)$ of H, we write $H - G = (V \setminus V_G, E_{H-G})$, the subgraph that contains all vertices that are not in V_G and all edges incident to $V \setminus V_G$.
We write $\Gamma(G)$ as the maximum of $\Delta(G)$ and $\Delta(H - G)$.

$$\Gamma(G) := \max\{\Delta(G), \Delta(H - G)\}. \tag{13.2.1}$$

Notation 13.18. We let Q_n denote a Boolean cube of dimension n.

Theorem 13.19 (Gotsman-Lineal Theorem [GL92]). *For any monotone function $\phi : \mathbb{N} \to \mathbb{R}$, the following are equivalent:*

(a) For any induced subgraph $G \subseteq Q_n$ with $|V(G)| \neq 2^{n-1}$, we have

$$\Gamma(G) \geq \phi(n) \tag{13.2.2}$$

(b) For any Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$,

$$s(f) \geq \phi(\deg(f)) \tag{13.2.3}$$

We first present two auxiliary lemmas.

Lemma 13.20. *Let $p : \{-1, 1\}^n \to \{-1, 1\}$ is the parity function. For any Boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$, we have*

$$\forall S \in [n], \hat{f}(S) = \widehat{fp}([n] \setminus S). \tag{13.2.4}$$

Proof. Note that

$$\hat{f}(S) = \langle f, \chi_S \rangle = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x) \tag{13.2.5}$$

On the other hand,

$$\widehat{fp}(T) = \langle fp, \chi_T \rangle = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)p(x)\chi_T(x) \tag{13.2.6}$$

Because $p(x)$ is the parity function, $p(x)\chi_T(x) = \chi_{[n]\setminus T}(x)$. Setting $S = [n]\setminus T$, we then have $\hat{f}(S) = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x) = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\chi_{[n]\setminus T}(x) = \widehat{fp}(T) = \widehat{fp}([n] \setminus S)$. □

Lemma 13.21. *Let $p : \{-1,1\}^n \to \{-1,1\}$ is the parity function. For any $f : \{-1,1\}^n \to \{-1,1\}$ with degree n and any induced subgraph G of Q_n such that*

$$V(G) = \{x \in \{-1,1\}^n | f(x) \cdot p(x) = 1\} \text{ where } p \text{ is the parity function} \tag{13.2.7}$$

we have $\Gamma(G) = s(f)$.

Proof. Note that because p is a parity function, which is sensitive to every bit, we must have $s(fp, x) = n - s(f, x)$. Hence, we have that $\forall x \in V(G)$,

$$\begin{aligned} \deg_G(x) &= n - s(fp, x) = s(f, x) \\ \deg_{Q_n - G}(x) &= s(f, x) \end{aligned} \tag{13.2.8}$$

In other words, we have $\Gamma(G) = s(f)$. □

Now we are ready to prove the Gotsman-Lineal Theorem.

proof of Gotsman-Lineal Theorem. We first show that (b) is equivalent to the following statement:

(b') For any Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ with degree n,

$$s(f) \geq \phi(n) \tag{13.2.9}$$

Notice that the $(b) \Rightarrow (b')$ is trivial. To show that $(b') \Rightarrow (b)$, we let f be a Boolean function whose degree $d < n$. Then $\hat{f}(S) \neq 0$ for some $S \subseteq [n], |S| = d$. Without loss of generality, suppose $x_1, ..., x_d$ appears with nonzero coefficient in the Fourier expansion of f; for example $f([d]) \neq 0$. Let $g : \{-1,1\}^d \to \{-1,1\}$ be defined as

$$g(x_1, x_2, ..., x_d) := f(x_1, x_2, ..., x_d, 0, ..., 0) \tag{13.2.10}$$

We then have by the assumption in (b'), $s(f) \geq s(g) \geq \phi(d) = \phi(\deg(f))$.

It remains to show that $(a) \Leftrightarrow (b')$.
We first show $(a) \Rightarrow (b')$ by contraposition. Suppose that $f : \{-1,1\}^n \to \{-1,1\}$ is a Boolean function with $\deg(f) = n$ and $s(f) < \phi(n)$. Let G be an induced subgraph G of Q_n such that

$$V(G) = \{x \in \{-1,1\}^n | f(x) \cdot p(x) = 1\} \text{ where } p \text{ is the parity function} \tag{13.2.11}$$

By lemma 13.20, we have that $\widehat{fp}(\emptyset) = \hat{f}([n])$, which is nonzero because $\deg(f) = n$. This implies that $|V(G)| \neq 2^{n-1}$. The reason is that if $|V(G)| = 2^{n-1}$, then $f(x)p(x) = 1$ for exactly half of the time and $f(x)p(x) = -1$ for exactly the other half of the time. However, $\mathbb{E}_{x \in \{0,1\}^n}[f(x)p(x)] = \widehat{fp}(\emptyset) \neq 0$. By lemma 13.21, we have $\Gamma(G) = s(f) < \phi(n)$, which leads to the contrapositive of (a).

We again show $(b') \Rightarrow (a)$ by contraposition. Let the induced subgraph $G \subseteq Q_n$ with $|V(G)| \neq 2^{n-1}$ satisfy $\Gamma(G) < \phi(n)$. We define a function $f : \{-1,1\}^n \to \{-1,1\}$ by

$$f(x)p(x) = 1 \Leftrightarrow x \in V(G) \tag{13.2.12}$$

Again, because $|V(G)| \neq 2^{n-1}, 0 \neq \mathbb{E}_{x \in \{0,1\}^n}[f(x)p(x)] = \widehat{fp}(\emptyset) = \hat{f}([n])$. Hence, f has degree n. Also, by lemma 13.21, $s(f) = \Gamma(G) < \phi(n)$, which leads to the contrapositive of (b'). □

13.3 Huang's result ❖

In [Hua19], Huang proved part (a) of the Gotsman-Lineal Theorem with $\phi(t) = \sqrt{t}$. Specifically,

Theorem 13.22 ([Hua19])**.** *For any induced subgraph $G \subseteq Q_n$ with $|V(G)| \neq 2^{n-1}$, we have*

$$\Gamma(G) \geq \sqrt{n} \tag{13.3.1}$$

Because for any Boolean function $f : \{-1,1\}^n \to \{-1,1\}, s(f) = \Gamma(G)$ by lemma 13.21 and $\deg(f) \leq n$, we immediately have the following corollary.

Corollary 13.23. *For any Boolean function $f : \{-1,1\}^n \to \{-1,1\}, \deg(f) \leq s(f)^2$.*

To prove, Huang's theorem, we will need a few auxiliary lemmas from spectral graph theory and linear algebra.

Lemma 13.24. *For any graph $G, \Delta(G) \geq |\lambda_1|$ where λ_1 is the largest eigenvalue of the adjacency matrix of G.*

Proof. Let $A \in \{1,0\}^{n\times n}$ be the adjacency matrix of G and let v be the eigenvector of A with respect to λ_1. Hence, we have $\lambda_1 v = Av$. Without loss of generality, we can assume $v_1 \geq v_j$ for all $1 < j \leq m$. Hence,

$$|\lambda_1 v_1| = |(Av)_1| = |\sum_{j=1}^{n} A_{1j}v_j| = |\sum_{j:A_{1j}=1} A_{1j}v_j| \leq |\sum_{j:A_{1j}=1} A_{1j}v_1| \leq \sum_{j:A_{1j}=1} |A_{1j}||v_1| \tag{13.3.2}$$

Because $\Delta(G) = \sum_{j:A_{1j}=1} |A_{1j}|$, we obtain that $|\lambda_1 v_1| \leq \Delta(G)|v_1|$ and thus $\lambda_1 \leq \Delta(G)$. □

Remark 13.25. Notice that the lemma remains true if some $+1$ in A is replaced by -1 because $|\sum_{j:A_{1j}=1} A_{1j}v_1| \leq \sum_{j:A_{1j}=1} |A_{1j}||v_1|$ still holds.

We state the following result from linear algebra without proof.

Theorem 13.26 (Cauchy's Interlace Theorem)**.** *Let A be a symmetric $n \times n$ matrix and B be a $m \times m$ principal submatrix of A for some $m < n$. IF the eigenvalues of A are $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ and the eigenvalues of B are $\mu_1, \mu_2, ..., \mu_m$, then for all $1 \leq i \leq m$, we have*

$$\lambda_i \geq \mu_i \geq \lambda_{i+n-m} \tag{13.3.3}$$

The following corollary is immediate from the Cauchy's Interlace Theorem.

Corollary 13.27. *Let A be a symmetric $2^n \times 2^n$ matrix and B be a $2^{n-1}+1 \times 2^{n-1}+1$ principal submatrix of A. Then $\mu_1 \geq \lambda_{1+2^n-(2^{n-1}+1)} = \lambda_{2^{n-1}}$.*

Lemma 13.28. *Let $B_n \in \{0,1\}^{2^n\times 2^n}$ be the adjacency matrix of Q_n. There is a matrix $A_n \in \{-1,0,1\}^{2^n\times 2^n}$ obtained by replacing some of the $+1$ entries in B_n by -1 such that exactly a half of the eigenvalues of A_n are $\sqrt{n}$ and the other half $-\sqrt{n}$.*

Proof. Let $I_n \in \{0,1\}^{2^n\times 2^n}$ denote the identity matrix. Notice that B_n can be recursively defined as follows:

- $B_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.
- $B_{n+1} := \begin{bmatrix} B_n & I_n \\ I_n & B_n \end{bmatrix}$.

Similarly, we define A by

- $A_1 := B_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.
- $A_{n+1} := \begin{bmatrix} A_n & I_n \\ I_n & -A_n \end{bmatrix}$.

We now show that $A_n^2 = nI_n$. As for the base case, we have

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{13.3.4}$$

For $n > 1$, assume that $A_n^2 = nI_n$, we have

$$\begin{bmatrix} A_n & I_n \\ I_n & -A_n \end{bmatrix}^2 = \begin{bmatrix} A_n^2 + I_n^2 & A_nI_n - A_nI_n \\ A_nI_n - A_nI_n & A_n^2 + I_n^2 \end{bmatrix} = \begin{bmatrix} (n+1)I_n & 0 \\ 0 & (n+1)I_n \end{bmatrix} = (n+1)I_{n+1}. \tag{13.3.5}$$

This implies A_n has one unique eigenvalue n. Because A_n and A_n^2 have exactly the same set of eigenvectors, we have that the eigenvalue λ_i of A is either $\sqrt{n}$ or $-\sqrt{n}$.
Note that for a real symmetric $N \times N$ matrix $Q = (q_{ij})$, the sum of all eigenvalues $\sum_{i=1}^N \lambda_i$ is equal to the trace $\mathrm{tr}(Q) = \sum_{i=1}^N q_{ii}$. Hence, it must be the case that exactly a half of the eigenvalues of A_n are $\sqrt{n}$ and the other half $-\sqrt{n}$ as $\sum_{i=1}^N a_{ii} = 0$. □

proof of theorem 13.22. Let $B_n \in \{0,1\}^{2^n \times 2^n}$ be the adjacency matrix of Q_n. Let A be constructed as described in the above lemma. Take any $2^{n-1} + 1$ vertices of Q_n and let C be the corresponding $(2^{n-1} + 1) \times (2^{n-1} + 1)$ submatrix of A. Hence, we obtain an subgraph $G = (V, E)$ of Q_n with $|V| = 2^{n-1} + 1$ and by corollary 13.27 $\Delta(G) \geq \mu_1(C) \geq \lambda_{2^{n-1}}(A) = \sqrt{n}$. □

13.4 Consequences of the Sensitivity Conjecture ❖

14 Pseudorandomness

Let U_n denote the uniform distribution of an n-bit string. We can think of a general probability distribution as a non-negative function $D : \{-1,1\}^n \to \mathbb{R}_{\geq 0}$ satisfying

- $\forall x \in \{-1,1\}^n, D(x) \geq 0$, and
- $\sum_{x\in\{-1,1\}^n} D(x) = 1$.

The notion of probability density function follows naturally.

Definition 14.1. A **(probability) density** function $\varphi_D : \{-1,1\}^n \to \mathbb{R}_{\geq 0}$ of a distribution D is a nonnegative function defined by

$$\varphi_D(x) = 2^n D(x) \tag{14.0.1}$$

Notice that $\mathbb{E}_{x\sim U_n}[\varphi_D(x)] = 1$. Observe that for any $g : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbb{E}_{y\sim D}[g(y)] = \frac{1}{2^n}\langle \varphi_D, g\rangle = \frac{1}{2^n}\mathbb{E}_{x\sim U_n}[\varphi(x)g(x)] \tag{14.0.2}$$

In particular, we have that for a nonempty set $S \subseteq [n]$, if we replace g by χ_S,

$$\hat{\varphi}_D(S) = \mathbb{E}_{x\sim U_n}[\varphi(x)\chi_S(x)] = \mathbb{E}_{y\sim D}[\chi_S(y)] \tag{14.0.3}$$

Example 14.2. We have that the density function of a uniform distribution U_n is simply a constant 1 function and

$$\begin{aligned} \hat{\varphi}_{U_n}(\emptyset) &= \mathbb{E}_{y\sim U_n}[\chi_\emptyset(y)] = 1 \\ \hat{\varphi}_{U_n}(S) &= \mathbb{E}_{y\sim U_n}[\chi_S(y)] = 0 \quad \text{for } S \neq \emptyset \end{aligned} \tag{14.0.4}$$

Definition 14.3. Let D be a distribution and $\mathcal{A}$ be a family of functions $f : \{-1,1\}^n \to \mathbb{R}$. We say that the distribution **ϵ-fools $\mathcal{A}$** if

$$|\mathbb{E}_{y\sim D}[f(y)] - \mathbb{E}_{x\sim U_n}[f(x)]| \leq \epsilon \tag{14.0.5}$$

for any $f \in \mathcal{A}$.

Definition 14.4. For a distribution $D : \{-1,1\}^n \to \mathbb{R}_{\geq 0}$, we say that $\varphi_D : \{-1,1\}^n \to \mathbb{R}_{\geq 0}$ is a **ϵ-biased density function** if

$$|\hat{\varphi}_D(S)| \leq \epsilon \tag{14.0.6}$$

for any $S \neq \emptyset$. We also call D an ϵ-biased distribution.

In other words, the Fourier coefficients of the density function of a small biased distribution is "similar" to those of the uniform distribution.

Remark 14.5. Notice that "Bias" has several meanings:

- A distribution D with an ϵ-biased density function is called an ϵ-biased distribution.
- We say that a p-biased distribution if each bit is independently chosen to be 1 with probability p and -1 with probability $(1-p)$.
- As for a function $f : \{-1,1\}^n \to \{-1,1\}$, we have $\text{bias}(f) = \mathbb{E}_{x\sim U_n}[f(x)]$.
- We say that a function $f : \{-1,1\}^n \to \mathbb{R}$ is ϵ-regular if $|\hat{(}f)(S)| \leq \epsilon$ for all nonempty subset $S \subseteq [n]$. Notice that this is same as ϵ-biased for density functions $\varphi_D : \{-1,1\}^n \to \mathbb{R}_{\geq 0}$.

Definition 14.6 (Pseudorandom Generators)**.** We say that a function $G : \{-1,1\}^r \to \{-1,1\}^n, r < n$, is an **$\epsilon$-biased pseudorandom generator(PRG)** if the output distribution $G(U_r)$ is an ϵ-biased distribution.

Definition 14.7. The **sample space** or **support** of a distribution $D : \{-1,1\}^n \to \mathbb{R}_{\geq 0}$ is defined by

$$\Sigma_D = \{x \in \{-1,1\}^n | D(x) \neq 0\}. \tag{14.0.7}$$

Remark 14.8. Notice that because the input of G is from $\{-1,1\}^r$, the support of $G(U_r)$ has at most 2^r strings.

Definition 14.9. We say that $f : \{-1,1\}^n \to \{-1,1\}$ is a **Bent function** if $\hat{f}(S) = \pm 2^{-n/2}$ for all $S \subseteq [n]$.

Example 14.10. The *inner product* mod 2 function $f : \{-1,1\}^n \to \{-1,1\}, n = 2t$ has Fourier coefficients $\pm 2^{-n/2}$ and is thus a Bent function. The 0/1 version of the *inner product* mod 2 function $f : \{-1,1\}^n \to \{0,1\}, n = 2t$ can be viewed as a density function and its corresponding distribution is $1/2^{n/2}$-biased. However, the sample space is $\{-1,1\}^n$ and thus very large.

Essentially, we would like an ϵ-biased distribution with small sample space. In the next section, we will show an efficient constructions of ϵ-biased distributions in [AGHP92]. Before that, we present the following example to show the equivalence of constructing small biased distributions and fooling parity functions.

Example 14.11. Notice that if $G : \{-1,1\}^r \to \{-1,1\}^n, r << n$, is an ϵ-biased pseudorandom generator, then the output distribution $G(U_r)$ fools any parity function because for any nonempty set $S \subseteq [n]$,

$$|\mathbb{E}_{y \sim G(U_r)}[\chi_S(y)] - \mathbb{E}_{x \sim U_n}[\chi_S(x)]| = |\mathbb{E}_{y \sim G(U_r)}[\chi_S(y)]| = \hat{\varphi}_{G(U_r)}(S) \leq \epsilon \tag{14.0.8}$$

The first equality comes from the fact that the expectation of any parity function of a nonempty set is zero. The second equality follows from equation 14.0.3 and the third from the definition 14.4.
Notice that the output distribution has a sample space $|\Sigma_{G(U_r)}| \leq 2^r$ as shown in remark 14.8.

14.1 Efficient Constructions of ϵ-Biased Distributions with Small Sample Spaces ❖

In this section, we show the construction due to Alon, Goldreich, Hastad and Peralta that proves the following theorem.

Theorem 14.12 ([AGHP92])**.** *There exists a deterministic algorithm that given input n, ϵ, outputs in poly(n/ϵ) time a set $\Sigma \in \{0,1\}^n$ with $|\Sigma| = O((n/\epsilon)^2)$ such that the uniform distribution on Σ is ϵ-biased.*

This theorem implies that there is an ϵ-biased pseudorandom generator that maps $2\log(n/\epsilon)$ truly random bits to n bits. We remark that $|\Sigma| = O((n/\epsilon)^2)$ *is almost best possible.* Before we get into the proof, we introduced some concepts related to *finite fields.*

Proof of Theorem 14.12. Let $m = \log(n/\epsilon)$. Assume that an irreducible polynomial p of degree m is given. We use p to represent field elements in $\mathbb{F}_{2^m}$ as bit strings of length m.
Let (x, y) be an ordered pair where x and y are both of length m. We write x^i as the i-th power of x in $\mathbb{F}_{2^m}$ and $(x^i, y)_2$ the inner product of x^i and y, modular 2. Let $r = r_0 ... r_{n-1}$ be elements of the sample space Σ, where each

$$r_i = (x^i, y)_2. \tag{14.1.1}$$

Hence, we have

$$|\Sigma| = 2^m \times 2^m = (n/\epsilon)^2. \tag{14.1.2}$$

As we have seen in example 14.11, it suffices to show that the uniform distribution on $\Sigma\epsilon$-fools parity function χ_S for any $S \neq \emptyset$. Consider $s = s_0 ... s_{n-1}$, the characteristic vector of a nonempty set $S \subset [n]$, and the polynomial $p_s(t) = \sum_{i=0}^{n-1} s_i t^i$ over $\mathbb{F}_{2^m}$. Notice that because $r, s \in \{0,1\}^n$, the parity

$$\chi_S(r) = (-1)^{(s,r)_2} \tag{14.1.3}$$

Thus, it suffices to show that $(s, r)_2 = 1$ for $1/2 + \pm\epsilon$ fractions of $r \in \Sigma$. Now,

$$(s,r)_2 = \sum_{i=0}^{n-1} s_i r_i = \sum_{i=0}^{n-1} s_i (x^i, y)_2 = \sum_{i=0}^{n-1} s_i (\sum_{j=0}^{m-1} x_j^i \cdot y_j) = \sum_{j=0}^{m-1} (\sum_{i=0}^{n-1} s_i \cdot x_j^i) \cdot y_j = \sum_{j=0}^{m-1} (\sum_{i=0}^{n-1} s_i \cdot x^i)_j \cdot y_j \tag{14.1.4}$$

Because $p_s(t) = \sum_{i=0}^{n-1} s_i t^i$, we have

$$(s,r)_2 = \sum_{j=0}^{m-1} (p_s(x))_j \cdot y_j = (p_s(x), y)_2 \tag{14.1.5}$$

Consider all possible $y \in \{0,1\}^m$,

Case 1: If $p_s(x) \neq 0$ for fixed x, then $p_s(x)$ is also fixed. Therefore, $(p_s(x), y)_2 = 0$ for exactly half of all possible y because .

Case 2: If $p_s(x) = 0$, then $(p_s(x), y)_2 = 0$ for all y. However, because p_s has degree $n-1$, $p_s(x) = 0$ for at most $n-1$ possible x.

Overall, we have $(p_s(x), y)_2 = 0$ on at most

$$\frac{1}{2} + \frac{n-1}{2^m} < \frac{1}{2} + \frac{n}{2^m} = \frac{1}{2} + \epsilon$$

fraction of $(x,y) \in \{0,1\}^{2m}$ and thus at most $\frac{1}{2} + \epsilon$ fraction of $r \in \Sigma$. □

We now show that $|\Sigma| = O((n/\epsilon)^2)$ is almost optimal. Before that, we state Alon's result for perturbed identity matrices.

Theorem 14.13 ([Alo09]). *Let A be an $N \times N$ real matrix with $A_{i,i} = 1$ and $|A_{i,j}| \leq \epsilon$ for $i \neq j$, where $1/\sqrt{n} \leq \epsilon \leq 1/2$, then*

$$\operatorname{rank}(A) \geq \Omega(\frac{\log N}{\epsilon^2 \log(1/\epsilon)}) \tag{14.1.6}$$

Claim 14.14. The number of random bits $\Omega(\log(n/\epsilon))$ is optimal up to some hidden constant.

Proof. Let $G : \{0,1\}^r \to \{0,1\}^n$ be a ϵ-biased PRG. For a characteristic vector of a nonempty set $S \subset [n], s = s_0 ... s_{n-1}$, we define a real vector v_s of length 2^r by

$$v_s = \frac{1}{2^{r/2}} \Big((-1)^{(G(x),s)_2} : x \in \{0,1\}^r \Big) \tag{14.1.7}$$

Now, let A be a $2^n \times 2^n$ matrix defined by

$$A_{s,t} = (v_s, v_t) = \frac{1}{2^r} \sum_{x \in \{0,1\}^r} (-1)^{(G(x),s)_2} (-1)^{(G(x),t)_2} = \frac{1}{2^r} \sum_{x \in \{0,1\}^r} (-1)^{(G(x), s \oplus t)_2} \tag{14.1.8}$$

In other words, let V be a $2^n \times 2^r$ matrix with rows v_s, we define A by

$$A = V \cdot V^T \tag{14.1.9}$$

Because for any nonempty set $S, (G(x), s)_2 = 0$ for $1/2 \pm \epsilon$ fraction of time, we have

$$\begin{aligned} A_{s,s} &= 1 \\ |A_{s,t}| &\leq \epsilon \end{aligned} \tag{14.1.10}$$

Thus, by theorem 14.13, we have $\operatorname{rank}(A) \geq \Omega(\frac{\log N}{\epsilon^2 \log(1/\epsilon)})$. Because A can be at most full rank, i.e. $\operatorname{rank}(A) \leq 2^r$, we have

$$2^r \geq \Omega(\frac{\log N}{\epsilon^2 \log(1/\epsilon)}) \tag{14.1.11}$$

and thus $r \geq \Omega(\log(n/\epsilon))$. □

14.2 Higher Degree Polynomials ❖

So far, we have seen that ϵ-biased distributions can fool parity functions [NN93], [AGHP92], which are linear functions over any field. How about higher degree polynomials? The following example shows that small biased distributions may not fool quadratic polynomials.

Example 14.15. Consider the inner product mod 2 as a degree 2 polynomial over $\mathbb{F}_2$. That is, we define IP: $\mathbb{F}_2^n \to \mathbb{F}_2$ by

$$\mathrm{IP}(x) = x_1x_2 + x_3x_4 + ... + x_{n-1}x_n \tag{14.2.1}$$

We have seen in example 14.10, IP is an $1/2^{n/2}$-biased Bent function. Now, consider the $1/2^{n/2}$-biased distribution D with density function

$$\varphi_D = \mathrm{IP} \tag{14.2.2}$$

then we have the function IP is *constant* on the support of D and hence D does not fool the quadratic polynomial IP.

In [LVW93], Luby, Velickovic and Wigderson gave the first non-trivial PRG that fools *constant degree polynomials* and even *constant depth circuits*, but it requires *a large number of random bits.* In [Bog05], Bogdanov showed optimal constructions of PRGs but it only works when the field size is at least polynomial in the degree. In [BV07] Bogdanov and Viola gave a new approach, which we will present shortly, for polynomials in small fields; at the time they showed that the constructed PRG fooled polynomials of degree 2 and 3 and conjectured that it applied to any degree. Later in [Vio09], Viola proved this conjecture. Note also that in [Lov08], Lovett has shown a weaker version of this conjecture.

Theorem 14.16 ([Vio09])**.** *Fix $d \geq 1, \epsilon > 0$. Let D be an ϵ-biased distribution in $\{0,1\}^n$ and let $y^1, ..., y^d$ be independent samples from D. Then for any polynomial $h : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most d, then*

$$|\mathbb{E}_{y^1,...,y^d \sim D}[f(y^1 + ... + y^d)] - \mathbb{E}_{x \in U_n}[f(x)]| \leq \epsilon_d \tag{14.2.3}$$

where $f(x) = (-1)^{h(x)}, \epsilon_d = 16 \cdot \epsilon^{1/2^{d-1}}$.

To prove this theorem, we will need the following auxiliary lemma.

Lemma 14.17. *Let $h : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree d. We fix $y, y' \in \mathbb{F}_2^n$ and define $g : \mathbb{F}_2^n \to \mathbb{F}_2$ by*

$$g(x) := h(x + y) + h(x + y') \tag{14.2.4}$$

Then g has degree at most $d - 1$.

Sometimes we informally call g the derivative of h. For intuition, we can assume h and g are multilinear, but such an assumption is not needed for the proof.

Proof of lemma 14.17. Let $x^S, |S| = d$ be the monomial of $h(x)$ with the highest degree. We then have for $g : \mathbb{F}_2^n \to \mathbb{F}_2$, because y and y' are fixed, the term $(x + y)^S + (x + y')^S$ cancels out. Hence, g has degree at most $d - 1$. □

Notation 14.18. For the simplicity of notation, we write $z \sim D_{d-1}$ to denote $z = y^1 + ... + y^{d-1}$ for $y^1, ..., y^{d-1}$ sampled independently from D.

Proof of theorem 14.16. We prove the theorem by induction on the degree d.
Base Step: $d = 1$. By definition, because D is an ϵ-biased distribution, we have

$$|\mathbb{E}_{y \sim D}[f(y)] - \mathbb{E}_{x \in U_n}[f(x)]| \leq \epsilon \leq 16\epsilon = \epsilon_1. \tag{14.2.5}$$

Inductive Step: We assume that the statement holds for all degrees at most $d - 1$. Let $\delta = \sqrt{\epsilon_{d-1}}$, we split our analysis into two cases: $\mathbb{E}_{x \in U_n}[f(x)] > \delta$ and $\mathbb{E}_{x \in U_n}[f(x)] \leq \delta$.
Case 1: $\mathbb{E}_{u \sim U_n}[f(u)] > \delta$. We show that

$$|\mathbb{E}_{z \sim D_{d-1}}[f(z)] - \mathbb{E}_{x \sim U_n}[f(x)]| \leq \frac{\epsilon_{d-1}}{\delta} \tag{14.2.6}$$

We choose $z \sim D_{d-1}, x \sim U_n, y \sim U_n$ and set $y' = 0$, we write $g_y(x) = h(x+y) + h(x+y') = h(x+y) + h(x)$. Then,

$$\mathbb{E}_{y,z}[(-1)^{g_y(z)}] - \mathbb{E}_{y,x}[(-1)^{g_y(x)}] = \mathbb{E}_{y,z}[(-1)^{h(y+z)+h(z)}] - \mathbb{E}_{y,x}[(-1)^{h(y+x)+h(x)}] \tag{14.2.7}$$

Because y is uniformly random, $y+z$ and $y+z$ are also uniformly random. That is,

$$\begin{aligned}\mathbb{E}_{y,z}[(-1)^{g_y(z)}] - \mathbb{E}_{y,x}[(-1)^{g_y(x)}] &= \mathbb{E}_{y,z}[(-1)^{h(y)+h(z)}] - \mathbb{E}_{y,x}[(-1)^{h(y)+h(x)}] \\ &= \mathbb{E}_y[f(y)] \cdot (\mathbb{E}_z[f(z)] - \mathbb{E}_x[f(x)])\end{aligned} \tag{14.2.8}$$

On the other hand, by lemma 14.17, for fixed $y \in \mathbb{F}_2^n$ and $y' = 0$, $g_y(x)$ has degree at most $d-1$. By the inductive hypothesis, $|\mathbb{E}_z[(-1)^{g_y(z)}] - \mathbb{E}_x[(-1)^{g_y(x)}]| \leq \epsilon_{d-1}$ and thus

$$|\mathbb{E}_{y,z}[(-1)^{g_y(z)}] - \mathbb{E}_{y,x}[(-1)^{g_y(x)}]| \leq \mathbb{E}_y\Big[|\mathbb{E}_z[(-1)^{g_y(z)}] - \mathbb{E}_x[(-1)^{g_y(x)}]|\Big] \leq \epsilon_{d-1} \tag{14.2.9}$$

Recall that $\mathbb{E}_{y\sim U_n}[f(y)] > \delta$, which implies

$$|\mathbb{E}_z[f(z)] - \mathbb{E}_x[f(x)]| = \frac{|\mathbb{E}_{y,z}[(-1)^{g_y(z)}] - \mathbb{E}_{y,x}[(-1)^{g_y(x)}]|}{\mathbb{E}_y[f(y)]} \leq \frac{\epsilon_{d-1}}{\delta} \tag{14.2.10}$$

Now, let $z' \sim D_d, w \sim D_{d-1}, v \sim D$. Notice that since $x \sim U_n$, $v+x$ is uniformly random. Hence,

$$\begin{aligned}|\mathbb{E}_{z'\sim D_d}[f(z)] - \mathbb{E}_x[f(x)]| &= |\mathbb{E}_{v,w}[f(v+w)] - \mathbb{E}_{v,x}[f(v+x)]| \\ &\leq \mathbb{E}_v|\mathbb{E}_w[f(v+w)] - \mathbb{E}_x[f(v+x)]|\end{aligned} \tag{14.2.11}$$

Notice that for fixed v, $v + D_{d-1}$ is similar to $v + U_n$ to the polynomial f, i.e., $v + D_{d-1}$ also fools f. To see this, suppose on the contrary that $v + D_{d-1}$ does not fool f, then for the polynomial $f'(x) = f(v+x)$, we have

$$\epsilon_{d-1} < |\mathbb{E}_w[f(v+w)] - \mathbb{E}_x[f(v+x)]| = |\mathbb{E}_w[f'(w)] - \mathbb{E}_x[f'(x)]| \tag{14.2.12}$$

That is, D_{d-1} also fools f'. Because for fixed v, $f'(x)$ has the same degree as $f(v+x)$, this contradicts the inductive hypothesis. Now, because for fixed v, we have $v+x$ is also uniformly distributed and hence $\mathbb{E}_x[f(v+x)] > \delta$. This gives us $|\mathbb{E}_w[f(v+w)] - \mathbb{E}_x[f(v+x)]| \leq \epsilon_{d-1}/\delta$ and as $\delta = \sqrt{\epsilon_{d-1}}$,

$$|\mathbb{E}_{z'\sim D_d}[f(z)] - \mathbb{E}_x[f(x)]| \leq \mathbb{E}_v|\mathbb{E}_w[f(v+w)] - \mathbb{E}_x[f(v+x)]| \leq \frac{\epsilon_{d-1}}{\delta} = \sqrt{\epsilon_{d-1}} \leq \epsilon_d \tag{14.2.13}$$

Case 2: $\mathbb{E}_{u\sim U_n}[f(u)] \leq \delta$. We show that

$$|\mathbb{E}_{z'\sim D_d}[f(z')] - \mathbb{E}_x[f(x)]| \leq O(\delta + \sqrt{\epsilon_{d-1}}) \tag{14.2.14}$$

Let $x \sim U_n, v \sim D, v' \sim D, w \sim D_{d-1}, z' \sim D_d$. Notice

$$|\mathbb{E}_{z'}[f(z')]|^2 = |\mathbb{E}_{w,v}[f(w+v)]|^2 \tag{14.2.15}$$

Using Cauchy-Schwarz inequality, we have

$$|\mathbb{E}_{w,v}[f(w+v)]|^2 \leq \mathbb{E}_w\Big[|\mathbb{E}_v[f(w+v)]|^2\Big] \tag{14.2.16}$$

Hence,

$$\begin{aligned}|\mathbb{E}_{z'}[f(z')]|^2 \leq \mathbb{E}_w\Big[|\mathbb{E}_v[f(w+v)]|^2\Big] &= \mathbb{E}_w\Big[\mathbb{E}_{v,v'}[f(w+v)f(w+v')]\Big] \\ &= \mathbb{E}_{v,v'}\Big[\mathbb{E}_w[f(w+v)f(w+v')]\Big]\end{aligned} \tag{14.2.17}$$

Notice that for fixed v and v', we have

$$f(w+v)f(w+v') = (-1)^{h(w+v)+h(w+v')} \tag{14.2.18}$$

By lemma 14.17, $g_{v,v'}(w) = h(w+v) + h(w+v')$ has degree at most $d-1$. Hence, by the inductive hypothesis, let $f'(w) = f(w+v)f(w+v')$, we have

$$\begin{aligned}\mathbb{E}_w[f(w+v)f(w+v')] - \mathbb{E}_x[f(x+v)f(x+v')] &\leq |\mathbb{E}_w[f(w+v)f(w+v')] - \mathbb{E}_x[f(x+v)f(x+v')]| \\ &\leq \epsilon_{d-1}\end{aligned} \tag{14.2.19}$$

Hence,

$$|\mathbb{E}_{z'}[f(z')]|^2 \leq= \mathbb{E}_{v,v'}\Big[\mathbb{E}_w[f(w+v)f(w+v')]\Big] \leq \epsilon_{d-1} + \mathbb{E}_{v,v'}\Big[\mathbb{E}_x[f(x+v)f(x+v')]\Big] \tag{14.2.20}$$

Using Fourier expansion, we obtain,

$$\begin{aligned}\mathbb{E}_{v,v'}\Big[\mathbb{E}_x[f(x+v)f(x+v')]\Big] &= \mathbb{E}_{v,v'}\Big[\mathbb{E}_x[\Big(\sum_{S\subseteq[n]}\hat{f}(S)\chi_S(x+v)\Big)\Big(\sum_{T\subseteq[n]}\hat{f}(T)\chi_T(x+v')\Big)]\Big] \\ &= \mathbb{E}_{v,v'}\Big[\mathbb{E}_x[\sum_{S,T\subseteq[n]}\hat{f}(S)\hat{f}(T)\chi_S(x+v)\chi_T(x+v')]\Big] \\ &= \mathbb{E}_{v,v'}\Big[\mathbb{E}_x[\sum_{S,T\subseteq[n]}\hat{f}(S)\hat{f}(T)\chi_{S\Delta T}(x)\chi_S(v)\chi_T(v')]\Big]\end{aligned} \tag{14.2.21}$$

where $S\Delta T$ is the symmetric difference. By linearity of expectation, we have

$$\mathbb{E}_{v,v'}\Big[\mathbb{E}_x[f(x+v)f(x+v')]\Big] = \sum_{S,T\subseteq[n]}\hat{f}(S)\hat{f}(T)\mathbb{E}_{v,v'}[\chi_S(v)\chi_T(v')]\mathbb{E}_x[\chi_{S\Delta T}(x)] \tag{14.2.22}$$

By lemma 2.9, we have $\mathbb{E}_x[\chi_{S\Delta T}(x)] = 1$ if and only if $S = T$. Hence,

$$\begin{aligned}\mathbb{E}_{v,v'}\Big[\mathbb{E}_x[f(x+v)f(x+v')]\Big] &= \sum_{S\subseteq[n]}\hat{f}(S)^2\mathbb{E}_{v,v'}[\chi_S(v)\chi_S(v')] \\ &= \hat{f}(\emptyset)^2 + \sum_{|S|>0,S\subseteq[n]}\hat{f}(S)^2\mathbb{E}_{v,v'}[\chi_S(v)\chi_S(v')]\end{aligned} \tag{14.2.23}$$

By lemma 10.20, $\hat{f}(\emptyset) = \mathbb{E}_{u\sim U_n}[f(u)] \leq \delta$. Again, by Cauchy-Schwarz inequality, $\mathbb{E}_{v,v'}[\chi_S(v)\chi_S(v')] \leq |\mathbb{E}_v[\chi_S(v)]\mathbb{E}_{v'}[\chi_S(v')]|$. Notice that by equation 14.0.3, for any set S, $\mathbb{E}_{v'}[\chi_S(v')] = \phi_D(S) \leq \epsilon$ because D is ϵ-biased. In summary, we have

$$\begin{aligned}|\mathbb{E}_{z'}[f(z')]|^2 &\leq \epsilon_{d-1} + \mathbb{E}_{v,v'}\Big[\mathbb{E}_x[f(x+v)f(x+v')]\Big] \\ &\epsilon_{d-1} + \hat{f}(\emptyset)^2 + \sum_{|S|>0,S\subseteq[n]}\hat{f}(S)^2\mathbb{E}_{v,v'}[\chi_S(v)\chi_S(v')] \\ &= \epsilon_{d-1} + \delta^2 + \epsilon^2\sum_{|S|>0,S\subseteq[n]}\hat{f}(S)^2 \leq \epsilon_{d-1} + \delta^2 + \epsilon^2\end{aligned} \tag{14.2.24}$$

Because $\delta = \sqrt{\epsilon_{d-1}} = \epsilon_d$, we have $|\mathbb{E}_{z'}[f(z')]|^2 \leq 2\epsilon_{d-1} + \epsilon^2 \leq 4\epsilon_{d-1}$. Lastly, we have

$$|\mathbb{E}_{z'}[f(z')] - \mathbb{E}_{x\sim U_n}[f(x)]| \leq |\mathbb{E}_{z'}[f(z')]| + |\mathbb{E}_{x\sim U_n}[f(x)]| \leq \sqrt{4\epsilon_{d-1}} + \delta =\leq 3\epsilon_d \tag{14.2.25}$$

□

14.3 Applications ❖

ϵ-biased pseudorandom generators can be used to derandomise any randomised algorithm that is based on the *properties of parity functions over the uniform distribution.* In particular, we can use it to derandomise learning and testing algorithms, such as the BLR test (section 4), the Goldreich-Levin algorithm (section 5.2) and learning algorithms based on the Goldreich-Levin algorithm.

15 Noise Operator

Definition 15.1. For $0 \geq \rho \geq 1$ and $f : \{-1,1\}^n \to \mathbb{R}$, we define the **noise operator** with parameter $\rho, T_\rho(f) : \{-1,1\}^n \to \mathbb{R}$ on the function f by

$$T_\rho(f)(x) = \mathbb{E}_{y\sim\mu_\rho}[f(y)] \tag{15.0.1}$$

where we write $y \sim \mu_\rho$ to denote y as a "noisy copy" of x, i.e.,

$$y_i = \begin{cases} -x_i & \text{with probability } \frac{1}{2}(1-\rho) \\ x_i & \text{with probability } \frac{1}{2}(1+\rho) \end{cases}$$

In particular, we say that y is **ρ-correlated** to x.

Notice that $y = x$ when $\rho = 1$, which means $T_1(f)(x) = f(x)$ and y is uniformly random if $\rho = 0$. The noise operator is linear. That is, for some real number λ,

$$T_\rho(f + \lambda g) = T_\rho(f) + \lambda T_\rho(g) \tag{15.0.2}$$

Example 15.2. For fixed $x \in \{-1,1\}^n$, we have that for all i,

$$\mathbb{E}_{y\sim\mu_\rho}[y_i] = \rho x_i, \qquad \mathbb{E}_{y\sim\mu_\rho}[x_i y_i] = \rho \tag{15.0.3}$$

Proposition 15.3.

$$T_\rho(f) = \sum_{S\subseteq[n]} \rho^{|S|}\hat{f}(s)\chi_S \tag{15.0.4}$$

Proof.

$$T_\rho(f) = \mathbb{E}_{y\in\mu_\rho}[\chi_S(y)] = \mathbb{E}_{y\in\mu_\rho}[\prod_{i\in S} y_i] = \prod_{i\in S}\mathbb{E}_{y\in\mu_\rho}[y_i] = \prod_{i\in S}\rho x_i = \rho^{|S|}\hat{f}(s)\chi_S$$

□

Example 15.4. If $x \in \{-1,1\}^n$ is chosen uniformly at random and for each $x, y \sim \mu_\rho$, we say that (x,y) is a **ρ-correlated pair** of random strings. Notice that for all i,

$$\mathbb{E}[x_i] = \mathbb{E}[y_i] = 0, \qquad \mathbb{E}[x_i y_i] = \rho \tag{15.0.5}$$

15.1 L_p Norms ❖

Definition 15.5. Let V be a set. A function $\varphi : V \to \mathbb{R}$ is a **norm** if for all $u, v \in V$, we write $||v|| = \phi(v)$ and

1. $||v|| \geq 0$ and $||v|| = 0$ if and only if v is the identity element.
2. $||\alpha \cdot v|| = ||\alpha|| \cdot ||v||$ for any scalar $\alpha \in \mathbb{R}$.
3. $||u + v|| \leq ||u|| + ||v||$, which is the triangle inequality.

Recall that for $f : \{-1,1\}^n \to \mathbb{R}$, the L_p norm is defined as

$$||f||_p := \left(\mathbb{E}_{x\sim U_n}[|f(x)|^p]\right)^{1/p} \tag{15.1.1}$$

Notice that $||f||_p$ is a norm when $p \geq 1$; otherwise $||f||_p$ may violate the triangle inequality.

Lemma 15.6. *For any $c \in \mathbb{R}$ and $f : \{-1,1\}^n \to \mathbb{R}$,*

$$||cf||_p \leq |c|||f||_p \tag{15.1.2}$$

Definition 15.7. A real valued function f is **convex** if for any $0 \leq \lambda \leq 1$, we have

$$f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b) \tag{15.1.3}$$

Theorem 15.8 (Jensen's Inequality)**.** *If g is convex and $0 \leq \lambda_i \leq 1, \sum_i \lambda_i = 1$, then*

$$g(\sum_i \lambda_i x_i) \leq \sum_i \lambda_i g(x_i) \tag{15.1.4}$$

Theorem 15.9 (Minkowski's Inequality)**.** *Let $f, g : \{-1,1\}^n \to \mathbb{R}$,*

$$||f+g||_p \leq ||f||_p + ||g||_p. \tag{15.1.5}$$

Notice that for the vector norms, the opposite of the following result holds.

Proposition 15.10. *For $1 \leq p \leq q \leq \infty$, we have*

$$||f||_p \leq ||f||_q \tag{15.1.6}$$

Proof. We write $\pi(x)$ to denote the probability of drawing x from some distribution; as in the case for L_p norm, $\pi(x) = 2^{-n}$.

$$\mathbb{E}_{x \in U_n}[|f(x)|^p] = \sum_x p(x)|f(x)|^p \tag{15.1.7}$$

Let $p = 1$ and $q > 1$. Let $g(z) = z^{q/p} = z^q$, we have that g is convex and thus by Jensen's Inequality,

$$g(\sum_x p(x)|f(x)|) = \Big(\sum_x \pi(x)|f(x)|\Big)^q \leq \sum_x p(x)|f(x)|^q \tag{15.1.8}$$

Hence,

$$\Big(\mathbb{E}_{x \in U_n}[|f(x)|^p]\Big)^{1/p} = \Big(\sum_x \pi(x)|f(x)|\Big) \leq \Big(\sum_x \pi(x)|f(x)|^q\Big)^{1/q} = \Big(\mathbb{E}_{x \in U_n}[|f(x)|^q]\Big)^{1/q} \tag{15.1.9}$$

As for the general case, we have $g(z) = z^{q/p}$ is convex because $q \geq p$. Hence,

$$g(\sum_x p(x)|f(x)|^p) = \Big(\sum_x \pi(x)|f(x)|^p\Big)^{q/p} \leq \sum_x p(x)|f(x)|^q \tag{15.1.10}$$

□

15.2 Contractivity ❖

Let $x, y \in \{-1,1\}^n$ where y is ρ-correlated to x. It is sometimes convenient to think of y as the coordinate-wise product as $y = x \cdot z$ where

$$z_i = \begin{cases} -1 & \text{with probability } \frac{1}{2}(1-\rho) \\ 1 & \text{with probability } \frac{1}{2}(1+\rho) \end{cases} \tag{15.2.1}$$

Hence, we can write

$$T_\rho(f)(x) = \mathbb{E}_z[f(x \cdot z)] \tag{15.2.2}$$

Proposition 15.11. *The noise operator $T_\rho(f)$ with parameter $0 \leq \rho \leq 1$ on function $f : \{-1,1\}^n \to \mathbb{R}$ is* **contractive**, *i.e. for $p \geq 1$,*

$$||T_\rho(f)||_p \leq ||f||_p \tag{15.2.3}$$

Proof. Notice

$$\begin{aligned} ||T_\rho(f)||_p &= \left(\mathbb{E}_{x\sim U_n}\left[|\mathbb{E}_{y\sim\mu_\rho}[f(y)]|^p\right]\right)^{1/p} \\ &= \left(\mathbb{E}_{x\sim U_n}\left[|\mathbb{E}_z[f(x\cdot z)]|^p\right]\right)^{1/p} \\ &= \left(\mathbb{E}_{x\sim U_n}\left[\Big|\sum_z \pi(z)f(x\cdot z)\Big|^p\right]\right)^{1/p} \end{aligned} \tag{15.2.4}$$

For each z, let $g_z(x) = \pi(z)f(x\cdot z)$. By Minkowski's inequality, we have

$$\left(\mathbb{E}_{x\sim U_n}\left[\Big|\sum_z \pi(z)f(x\cdot z)\Big|^p\right]\right)^{1/p} = \left(\mathbb{E}_{x\sim U_n}\left[\Big|\sum_z g_z(x)\Big|^p\right]\right)^{1/p} = ||\sum_z g_z||_p \le \sum_z ||g_z||_p \tag{15.2.5}$$

Because for fixed z, $\pi(z)$ is fixed and $x\cdot z$ is uniformly random as x is uniform.

$$||g_z||_p = \left(\mathbb{E}_{x\sim U_n}[|\pi(z)f(x\cdot z)|^p]\right)^{1/p} = \left(\mathbb{E}_{x\sim U_n}[|\pi(z)f(x)|^p]\right)^{1/p} = ||\pi(z)f||_p \tag{15.2.6}$$

By lemma 15.6, $||\pi(z)f||_p = |\pi(z)|\cdot||f||_p$. In summary,

$$||T_\rho(f)||_p \le \sum_z ||g_z||_p = \sum_z \pi(z)\cdot||f||_p = \mathbb{E}_z[||f||_p] = ||f||_p \tag{15.2.7}$$

□

Remark 15.12. Notice that for $0 \le \rho < \rho' \le 1$, we have

$$T_\rho(f) = T_{\rho''}(T_{\rho'}(f)) \qquad \text{where } \rho'' = \frac{\rho}{\rho'} \tag{15.2.8}$$

Let $g = T_{\rho'}(f)$, using the theorem we just proved, we have

$$||T_{\rho'}(f)||_p = ||g||_p \ge ||T_{\rho''}(g)||_p = ||T_{\rho''}(T_{\rho'}(f))||_p = ||T_\rho(f)||_p \tag{15.2.9}$$

15.3 Hypercontractivity Theorem ❖

Recall that for $1 \le p \le q \le \infty$, we have $||f||_p \le ||f||_q$. Hence, for all r with $1 \le r \le p$, we have $||T_\rho(f)||_r \le ||T_\rho(f)||_p \le ||f||_p$. Is there any ρ such that for any $r > p$, $||T_\rho(f)||_r \le ||f||_p$ still holds? In fact, we can show the following theorem.

Theorem 15.13 (Hypercontractivity Theorem)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$. For any $1 \le p \le q \le \infty$ and*

$$0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$$

we have $||T_\rho(f)||_q \le ||f||_p$.

We prove this by induction on n.

15.3.1 Base step: $n = 1$

Let $a = f(1), b = f(-1)$. Without loss of generality, we assume that $a > 0, b > 0$.

$$\begin{aligned}
||T_\rho(f)||_q &= \Big(\mathbb{E}_{x\sim U_n}\Big[|\mathbb{E}_{y\sim\mu_\rho}[f(y)]|^q\Big]\Big)^{1/q} = \Big(\mathbb{E}_{x\sim U_n}\Big[|\mathbb{E}_z[f(x\cdot z)]|^q\Big]\Big)^{1/q} \\
&= \Big[\frac{1}{2}\Big|\frac{1}{2}(1-\rho)f(1) + \frac{1}{2}(1+\rho)f(-1)\Big|^q + \frac{1}{2}\Big|\frac{1}{2}(1+\rho)f(1) + \frac{1}{2}(1-\rho)f(-1)\Big|^q\Big]^{1/q} \\
&= \Big[\frac{1}{2}\Big|\frac{1}{2}(1-\rho)a + \frac{1}{2}(1+\rho)b\Big|^q + \frac{1}{2}\Big|\frac{1}{2}(1+\rho)a + \frac{1}{2}(1-\rho)b\Big|^q\Big]^{1/q} \\
&= \Big[\frac{1}{2}\Big(\frac{a+b}{2} - \rho\frac{a-b}{2}\Big)^q + \frac{1}{2}\Big(\frac{a+b}{2} + \rho\frac{a-b}{2}\Big)^q\Big]^{1/q} \\
&= \frac{a+b}{2}\Big[\frac{1}{2}\Big(1 - \rho\frac{a-b}{a+b}\Big)^q + \frac{1}{2}\Big(1 + \rho\frac{a-b}{a+b}\Big)^q\Big]^{1/q}
\end{aligned} \tag{15.3.1}$$

Let $\alpha = (a-b)/(a+b)$. Then

$$||T_\rho(f)||_q = \frac{a+b}{2}\Big[\frac{1}{2}\Big(1-\rho\alpha\Big)^q + \frac{1}{2}\Big(1+\rho\alpha\Big)^q\Big]^{1/q} \tag{15.3.2}$$

Notice

$$||f||_q = ||T_1(f)||_q = \frac{a+b}{2}\Big[\frac{1}{2}\Big(1-\alpha\Big)^q + \frac{1}{2}\Big(1+\alpha\Big)^q\Big]^{1/q} \tag{15.3.3}$$

Applying Taylor expansion, we have

$$\begin{aligned}
(1+\rho\alpha)^q &= 1 + \rho q\alpha + \frac{\rho^2}{2}q(q-1)\alpha^2 + ... \\
(1-\rho\alpha)^q &= 1 - \rho q\alpha + \frac{\rho^2}{2}q(q-1)\alpha^2 - ... \\
(1+\rho\alpha)^q + (1-\rho\alpha)^q &= 1 + \frac{\rho^2}{2}q(q-1)\alpha^2 + ... \\
(1+\alpha)^p + (1-\alpha)^p &= 1 + \frac{p(p-1)}{2}\alpha^2 + ...
\end{aligned} \tag{15.3.4}$$

Hence, suppose $|\rho| \le \sqrt{(p-1)/(q-1)}$, then

$$\frac{\rho^2}{2}q(q-1)\alpha^2 \le q(p-1)\alpha^2 \tag{15.3.5}$$

One can check that

$$\Big[(1+\rho\alpha)^q + (1-\rho\alpha)^q\Big]^{1/q} \le \Big[(1+\alpha)^p + (1-\alpha)^p\Big]^{1/p} \tag{15.3.6}$$

which concludes the base step.

15.3.2 Notations

Before we start the inductive step, we introduce some notations.

Definition 15.14. Let $\mu : X \to \mathbb{R}$ be a probability distribution on X and $\nu : Y \to \mathbb{R}$ a probability distribution on Y. Then $\mu \times \nu$ is a probability distribution on $X \times Y$ and

$$\mathbb{P}_{\mu\times\nu}[x, y] = \mu(x) \cdot \nu(y) \tag{15.3.7}$$

We call $\mu \times \nu$ the **product distribution**.

For example, the uniform distribution on $\{-1, 1\}^n$ and the distribution μ_ρ used for noise operator are product distributions. Recall the subfunctions of $f : X \times Y \to R$:

- For a fixed $x \in X$, $f_x : Y \to R$ is a subfunction restricted to Y.
- For a fixed $y \in Y$, $f_y : X \to R$ is a subfunction restricted to X.

Let $X \times Y = \{-1, 1\}^n$, $\mu : X \to \mathbb{R}$ be a distribution on X and $\nu : Y \to \mathbb{R}$ a probability on Y. We can think of $||f_x||_p$ as a function of x, i.e., we define $G_p(x)$ as

$$G_p(x) := ||f_x||_p = \Big(\mathbb{E}_{y\sim\nu}[|f_x(y)|^p]\Big)^{1/p}$$

Thus, we can write

$$||f||_p = ||G_p||_p = \Big(\mathbb{E}_{x\sim\mu}[|G_p(x)|^p]\Big)^{1/p} = \Big(\mathbb{E}_{x\sim\mu}\mathbb{E}_{y\sim\nu}[|f(x,y)|^p]\Big)^{1/p} \tag{15.3.8}$$

We write T_ρ^S the noise operator applied to the coordinates in subset S, i.e., for $f : X \times Y \to \mathbb{R}$, let u and v be defined as z earlier (equation 15.2.1), i.e. $x \cdot u \sim \mu_\rho, y \cdot v \sim \nu_\rho$; we can write $T_\rho(f)(x, y)$ as

$$T_\rho(f)(x, y) = \mathbb{E}_u\mathbb{E}_v[f(x \cdot u, y \cdot v)] \tag{15.3.9}$$

as for fixed $x \in X$, we write $T_\rho^Y(f)(x, y)$ as

$$T_\rho^Y(f)(x, y) = \mathbb{E}_v[f(x, y \cdot v)] \tag{15.3.10}$$

For fixed x, notice that the subfunction of $T_\rho(f)$

$$T_\rho^Y(f)_x(y) = \mathbb{E}_v[f_x(y \cdot v)] \tag{15.3.11}$$

is equivalent to $T_\rho^Y(f_x)$. We write $H_q : X \to \mathbb{R}$ as

$$H_q(x) = ||T_\rho^Y(f)_x||_q \tag{15.3.12}$$

15.3.3 Inductive step

By induction hypothesis, we have $H_q(x) = ||T_\rho^Y(f)_x||_q \le ||f_x||_p$. Thus,

$$||T_\rho(f)||_q = ||T_\rho^X T_\rho^Y(f)||_q = \Big(\mathbb{E}_x\mathbb{E}_y\big[\big|\mathbb{E}_u\mathbb{E}_v[f(x \cdot u, y \cdot v)]\big|^q\big]\Big)^{1/q} \tag{15.3.13}$$

Using Jensen's inequality,

$$\mathbb{E}_x\mathbb{E}_y\big[\big|\mathbb{E}_u\mathbb{E}_v[f(x \cdot u, y \cdot v)]\big|^q\big] \le \mathbb{E}_x\mathbb{E}_u\Big[\mathbb{E}_y\big[\big|\mathbb{E}_v[f(x \cdot u, y \cdot v)]\big|^q\big]\Big] \tag{15.3.14}$$

Note that for fixed x and u,

$$\mathbb{E}_v[f(x \cdot u, y \cdot v)] = \mathbb{E}_v[f_{x\cdot u}(y \cdot v)] = T_\rho^Y(f)_{x\cdot u}(y) \tag{15.3.15}$$

Hence,

$$\mathbb{E}_y\big[\big|\mathbb{E}_v[f(x \cdot u, y \cdot v)]\big|^q\big] = \mathbb{E}_y\big[\big|T_\rho^Y(f)_{x\cdot u}(y)\big|^q\big] = |H_q(x \cdot u)|^q \le ||f_{x\cdot u}||_p^q \tag{15.3.16}$$

which gives us

$$||T_\rho(f)||_q \le \Big(\mathbb{E}_x\mathbb{E}_u\Big[||f_{x\cdot u}||_p^q\Big]\Big)^{1/q} = \Big(\mathbb{E}_x\mathbb{E}_u\Big[\big(\mathbb{E}_y[|f_{x\cdot u}(y)|^p]\big)^{q/p}\Big]\Big)^{1/q} \tag{15.3.17}$$

We denote $h(z) := \mathbb{E}_y[|f_z(y)|^p]$. Then $||h||_{q/p} = \big(\mathbb{E}_z[|h(z)|^{q/p}]\big)^{p/q}$ and by Jensen's inequality, because $q > p$,

$$\begin{aligned} ||h||_{q/p} = \big(\mathbb{E}_z[|\mathbb{E}_y[|f_z(y)|^p]|^{q/p}]\big)^{p/q} &\le \big(\mathbb{E}_z[|\mathbb{E}_y[|f_z(y)|^{p\cdot q/p}]|]\big)^{p/q} \\ &= \big(\mathbb{E}_y[|\mathbb{E}_z[|f_y(z)|^{p\cdot q/p}]|]\big)^{p/q} = \mathbb{E}_y[|||f_y|^p||_{q/p}] \end{aligned} \tag{15.3.18}$$

Hence,

$$\left(\mathbb{E}_x\mathbb{E}_u\left[\left(\mathbb{E}_y[|f_{x\cdot u}(y)|^p]\right)^{q/p}\right]\right)^{1/q} = \left(\mathbb{E}_x\mathbb{E}_u\left[\left|h(x\cdot u)\right|^{q/p}\right]\right)^{1/q} = (||h||_{q/p})^{1/p} \leq \left(\mathbb{E}_y[|||f_y|^p||_{q/p}]\right)^{1/p} \tag{15.3.19}$$

Observe

$$|||f_y|^p||_{q/p} = \left(\mathbb{E}_x\mathbb{E}_u\left[\left|f(x\cdot u,y)\right|^{p\cdot(q/p)}\right]\right)^{p/q} = \left(\mathbb{E}_x\mathbb{E}_u\left[\left|f(x\cdot u,y)\right|^{q}\right]\right)^{(1/q)\cdot p} = ||T_\rho^X(f_y)||_q^p \leq ||f_y||_p^p \tag{15.3.20}$$

where the last inequality follows from the inductive hypothesis. In summary,

$$||T_\rho(f)||_q \leq \left(\mathbb{E}_x\mathbb{E}_u\left[\left(\mathbb{E}_y[|f_{x\cdot u}(y)|^p]\right)^{q/p}\right]\right)^{1/q} \leq \left(\mathbb{E}_y[||f_y||_p]\right)^{1/p} = ||f||_p \tag{15.3.21}$$

15.4 Matrix Valued Hypercontrativity Theorem ❖

Notice that when $q = 2$, by Parseval's Theorem 3.3 and Proposition 15.3, we have

$$||T_\rho(f)||_2 = (\mathbb{E}_{x\in U_n}[T_\rho(f)^2])^{1/2} = (\sum_{S\subseteq[n]} \rho^{2|S|}\hat{f}(S)^2)^{1/2} \tag{15.4.1}$$

Hence, for any $1 \leq p \leq 2$, by setting $\rho = \sqrt{p-1}$, we have by the Hypercontractivity Theorem 15.13,

$$(\sum_{S\subseteq[n]} \rho^{2|S|}\hat{f}(S)^2)^{1/2} = ||T_\rho(f)||_2 \leq ||f||_p = (\frac{1}{2^n}\sum_x |f(x)|^p)^{1/p} \tag{15.4.2}$$

In [BARDW08], Ben-Aroya, Regev and De Wolf presented a version of this inequality for matrix-valued functions on the Boolean cube.

Definition 15.15. Let M be a $d \times d$ complex matrix with singular values $\sigma_1, ..., \sigma_d$. The (normalised Schatten) p-norm is defined by

$$||M||_p = (\frac{1}{d}\sum_{i=1}^{d} \sigma_i^p)^{1/p} \tag{15.4.3}$$

Theorem 15.16 (Matrix Valued Hypercontrativity Theorem, [BARDW08])**.** *Let $\mathcal{M}$ be the space of $d\times d$ complex matrices. For any $f : \{-1,1\}^n \to \mathcal{M}$ and $1 \leq p \leq 2$, we have*

$$(\sum_{S\subseteq[n]} (\rho-1)^{|S|}\hat{f}(S)^2)^{1/2} \leq (\frac{1}{2^n}\sum_x |f(x)|^p)^{1/p} \tag{15.4.4}$$

In the same paper, they discussed applications of this theorem in one way quantum communication complexity and lower bounds on the length of 2-query locally decodable codes.

15.4.1 Locally Decodable Codes Revisited

Recall from section 5.1, we have seen that Hadamard code achieves optimal length for 2-query locally decodable codes. However, the 2^n-bit codewords are too large. On the other hand, for $q = 2$, we cannot do better than superpolynomial, as proved in [GKST02] for linear codes and later generalised in [KdW03].

Theorem 15.17 ([KdW03])**.** *For 2-query locally decodable codes, we have*

$$N \geq 2^{\Omega(n)} \tag{15.4.5}$$

Surprisingly, [KdW03] used quantum arguments to show this result. Later in [BARDW08], the same result was shown by applying the matrix valued hypercontrativity theorem.
As for 3-query locally decodable codes, Yekhanin has shown in [Yek08] that subexponential length 3-query linear locally decodable codes exist assuming infinitude of Mersenne primes, specifically

$$N = 2^{2^{O(\log n/\log\log n)}} \tag{15.4.6}$$

Note that this is smaller than 2^{n^ϵ} but larger than $n^{(\log n)^c}$. Later Woodruff [Woo12] gave an $\Omega(n^2)$ lower bound for linear 3-query locally decodable codes over any, possibly infinite, field, which is the largest general bound. In [GM12], Gal and Mills showed that even for 3-query locally decodable codes, N is exponential under some parameters.

15.5 Noise Stability ❖

Definition 15.18. For $f : \{-1,1\}^n \to \mathbb{R}$, let (x, y) be ρ-correlated pairs. The **noise stability** of f is defined as

$$\mathrm{Stab}_\rho(f) = \mathbb{E}_{(x,y)}[f(x)f(y)] \tag{15.5.1}$$

In particular, for $f : \{-1,1\}^n \to \{-1,1\}$, we have

$$\mathrm{Stab}_\rho(f) = 2\mathbb{P}_{(x,y)}[f(x) = f(y)] - 1 \tag{15.5.2}$$

Trivially, constant functions have stability 1. The Fourier expansion of noise stability is given by

$$\mathrm{Stab}_\rho(f) = \mathbb{E}_{(x,y)}[f(x)f(y)] = \mathbb{E}_{x\sim U_n}[f(x)T_{rho}(f)(x)] = \langle f, T_\rho(f)\rangle \tag{15.5.3}$$

Using Plancherel's Theorem 3.4, we obtain

$$\mathrm{Stab}_\rho(f) = \langle f, T_\rho(f)\rangle = \sum_{S\subseteq[n]} \hat{f}(S)\widehat{T_\rho(f)}(S) = \sum_{S\subseteq[n]} \hat{f}(S)^2\rho^{|S|} \tag{15.5.4}$$

This implies that for $f : \{-1,1\}^n \to \{-1,1\}$ that is unbiased, i.e., $\hat{f}(\emptyset) = 0$, we have

$$\mathrm{Stab}_\rho(f) = \sum_{S\subseteq[n]} \hat{f}(S)^2\rho^{|S|} = \sum_{S\neq\emptyset} \hat{f}(S)^2\rho^{|S|} \le \rho\sum_{S\neq\emptyset} \hat{f}(S)^2 \le \rho \tag{15.5.5}$$

Example 15.19. Dictator function is noise stable. $\mathrm{Stab}_\rho(x_i) = \rho$. Observe that this matches the lower bound of unbiased functions f. Specifically,

$$\mathrm{Stab}_\rho(f) = \sum_{S\neq\emptyset} \hat{f}(S)^2\rho^{|S|} = \rho \tag{15.5.6}$$

if and only if $\hat{f}(S) = 0$ for all S with $|S| > 1$.

In fact, we have the following proposition.

Proposition 15.20. *Let $f : \{-1,1\}^n \to \{-1,1\}$ with $\hat{f}(S) = 0$ for all $|S| > 1$. Then f is either constant, a dictator function x_i, or an anti-dictator function $-x_i$.*

If we require that the individual influence of a bit is not too large, i.e., $\mathtt{Inf}[f]$ is small, then majority function is the "stablest".

Example 15.21. Majority function is also noise stable.

$$\mathrm{Stab}_\rho(\mathtt{MAJ}) \sim \begin{cases} \frac{2}{n}\rho & \text{when } \rho \text{ is close to 0} \\ 1 - o(\sqrt{1-\rho} & \text{when } \rho \text{ is close to 1} \end{cases} \tag{15.5.7}$$

Definition 15.22. For $f : \{-1,1\}^n \to \mathbb{R}$, we first sample $x \sim U_n$ and obtain y from x by flipping each bit independently with probability $0 \geq \delta \leq 1$. The **noise sensitivity** of f is defined as

$$\mathrm{NS}_\delta(f) = \mathbb{P}[f(x) \neq f(y)] = \frac{1}{2} - \frac{1}{2}\mathrm{Stab}_{1-2\delta}(f) \tag{15.5.8}$$

Intuitively, f is "noise stable" if its value changes with small probability; on the other hand, f is "noise sensitive" if its value changes with large probability.

Example 15.23. Parity function is noise sensitive.

$$\begin{aligned} \mathrm{Stab}_\rho(\chi_{[n]}) &= \mathbb{E}_{(x,y)}[\prod_i x_i y_i] = \prod_i \mathbb{E}_{(x,y)}[x_i y_i] = \rho^n \\ \mathrm{NS}_\delta(\chi_{[n]}) &= \frac{1}{2} - \frac{1}{2}\mathrm{Stab}_{1-2\delta}(\chi_{[n]}) = \frac{1}{2}\big(1 - (1-2\delta)^n\big) \end{aligned} \tag{15.5.9}$$

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[AGM+13] Andris Ambainis, Yihan Gao, Jieming Mao, Xiaoming Sun, and Song Zuo. New upper bound on block sensitivity and certificate complexity in terms of sensitivity, 2013.

[AL93] Miklós Ajtai and Nathan Linial. The influence of large coalitions. *Combinatorica*, 13(2):129–145, 1993.

[Alo09] Noga Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability and Computing*, 18(1-2):3–15, 2009.

[AS11] Andris Ambainis and Xiaoming Sun. New separation between $s(f)$ and $bs(f)$. *arXiv preprint arXiv:1108.3494*, 2011.

[AUY83] Alfred V Aho, Jeffrey D Ullman, and Mihalis Yannakakis. On notions of information transfer in vlsi circuits. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 133–139, 1983.

[BARDW08] Avraham Ben-Aroya, Oded Regev, and Ronald De Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and ldcs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 477–486. IEEE, 2008.

[BI87] Manuel Blum and Russell Impagliazzo. Generic oracles and oracle classes. In *28th Annual Symposium on Foundations of Computer Science (sfcs 1987)*, pages 118–126. IEEE, 1987.

[Bog05] Andrej Bogdanov. Pseudorandom generators for low degree polynomials. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 21–30, 2005.

[BV07] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '07, pages 41–51, USA, 2007. IEEE Computer Society.

[CD82] Stephen Cook and Cynthia Dwork. Bounds on the time for parallel ram's to compute simple functions. In *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pages 231–233, 1982.

[CG18] Siddhesh Chaubal and Anna Gál. New constructions with quadratic separation between sensitivity and block sensitivity. In *38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[Cha05] Sourav Chakraborty. On the sensitivity of cyclically-invariant boolean functions. In *20th Annual IEEE Conference on Computational Complexity (CCC'05)*, pages 163–167. IEEE, 2005.

[GKST02] Oded Goldreich, Howard Karloff, Leonard J Schulman, and Luca Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings 17th IEEE Annual Conference on Computational Complexity*, pages 175–183. IEEE, 2002.

[GL92] C Gotsman and N Linial. The equivalence of two problems on the cube. *Journal of Combinatorial Theory, Series A*, 61(1):142 – 146, 1992.

[GM12] Anna Gal and Andrew Mills. Three-query locally decodable codes with higher correctness require exponential length. *ACM Trans. Comput. Theory*, 3(2), January 2012.

[GRS12] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. *Draft available at http://www. cse. buffalo. edu/ atri/courses/coding-theory/book*, 2012.

[GSTW16] Parikshit Gopalan, Rocco Servedio, Avishay Tal, and Avi Wigderson. Degree and sensitivity: tails of two distributions. *arXiv preprint arXiv:1604.07432*, 2016.

[Har64] L. H. Harper. Optimal assignments of numbers to vertices. *Journal of the Society for Industrial and Applied Mathematics*, 12(1):131–135, 1964.

[Har66] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, 1(3):385 – 393, 1966.

[Hua19] Hao Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *Annals of Mathematics*, 190(3):949–955, 2019.

[KdW03] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 106–115, New York, NY, USA, 2003. Association for Computing Machinery.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *[Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science*, pages 68–80, Oct 1988.

[KM93] Eyal. Kushilevitz and Yishay. Mansour. Learning decision trees using the fourier spectrum. *SIAM Journal on Computing*, 22(6):1331–1348, 1993.

[LMN89] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform, and learnability. In *30th Annual Symposium on Foundations of Computer Science*, pages 574–579, Oct 1989.

[LNS20] Sophie Laplante, Reza Naserasr, and Anupa Sunny. Sensitivity lower bounds from linear dependencies. 2020.

[Lov08] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 557–562, New York, NY, USA, 2008. Association for Computing Machinery.

[LVW93] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *[1993] The 2nd Israel Symposium on Theory and Computing Systems*, pages 18–24. IEEE, 1993.

[Nis89] N. Nisan. Crew prams and decision trees. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, STOC '89, pages 327–335, New York, NY, USA, 1989. Association for Computing Machinery.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[NS92] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 462–467, 1992.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[OSSS05] R. O'Donnell, M. Saks, O. Schramm, and R. A. Servedio. Every decision tree has an influential variable. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 31–39, Oct 2005.

[Rei82] Rüdiger Reischuk. *A lower time-bound for parallel random access machines without simultaneous writes*. IBM Thomas J. Watson Research Center, 1982.

[Rub95] David Rubinstein. Sensitivity vs. block sensitivity of boolean functions. *Combinatorica*, 15(2):297–299, 1995.

[Tal94] Michel Talagrand. On russo's approximate zero-one law. *The Annals of Probability*, 22(3):1576–1587, 1994.

[Tal13] Avishay Tal. Properties and applications of boolean function composition. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 441–454, 2013.

[Tar89] Gábor Tardos. Query complexity, or why is it difficult to separate $np^a \cap conp^a$ from p^a by random oracles a? *Combinatorica*, 9(4):385–392, 1989.

[Vio09] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *computational complexity*, 18(2):209–217, 2009.

[Vir11] Madars Virza. Sensitivity versus block sensitivity of boolean functions. *Information Processing Letters*, 111(9):433–435, 2011.

[Woo12] David P Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. *Journal of Computer Science and Technology*, 27(4):678–686, 2012.

[Y+12] Sergey Yekhanin et al. Locally decodable codes. *Foundations and Trends® in Theoretical Computer Science*, 6(3):139–255, 2012.

[Yek08] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *J. ACM*, 55(1), February 2008.