Purdue University

CS 699

Spring 2016

Research Thesis

Author:
Hai Nguyen

Instructor:
Prof. Hemanta Maji

April 6, 2016

Chapter 1

Fourier Basics

1.1 Vector Space of Functions on Boolean Hyper-cube

Definition 1.1 (Inner Product)**.** Consider the 2^n-dimensional vector space of all functions $f : \{0,1\}^n \to \mathbb{R}$. We define an inner product on this space by

$$\langle f, g\rangle := \mathbb{E}[f \cdot g] = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} f(x)g(x)$$

.

1.2 Characteristic Functions

Definition 1.2 (Characteristic function)**.** For each $S \subseteq [n] = \{1, 2, ..., n\}$, we define the characteristic function of S as

$$\chi_S(x) = (-1)^{S\cdot x}, \text{where } S \cdot x = \sum_{i=1}^{n} S_i \cdot x_i = \sum_{i\in S} x_i$$

.

Lemma 1.3. For every $S \subseteq [n]$,

$$\sum_{x\in\{0,1\}^n} \chi_S(x) = \begin{cases} 2^n & \text{if } S = \emptyset \\ 0 & \text{if } S \neq \emptyset \end{cases}$$

Proof. If $S = \emptyset$, then $S \cdot x = 0$. So $\sum_{x\in\{0,1\}^n} \chi_S(x) = \sum_{x\in\{0,1\}^n} 1 = 2^n$.
If $S \neq \emptyset$, then there exists k such that $S_k \neq 0$. Hence,

$$\begin{aligned}
\sum_{x\in\{0,1\}^n} \chi_S(x) &= \sum_{x\in\{0,1\}^n} (-1)^{\sum\limits_{i\in S} x_i} \\
&= \sum_{x\in\{0,1\}^n} [(-1)^{x_k} \cdot (-1)^{\sum\limits_{i\in S\setminus\{k\}} x_i}] \\
&= \sum_{x_k\in\{0,1\}} (-1)^{x_k} \cdot \sum_{x\setminus x_k\in\{0,1\}^{n-1}} (-1)^{\sum\limits_{i\in S\setminus\{k\}} x_i} \\
&= \left[(-1)^0 + (-1)^1\right] \sum_{x\setminus x_k\in\{0,1\}^{n-1}} (-1)^{\sum\limits_{i\in S\setminus\{k\}} x_i} \\
&= 0
\end{aligned}$$

□

Theorem 1.4. *For every $S, T \subseteq [n]$,*

$$\langle \chi_S, \chi_T \rangle = \begin{cases} 1 & \textit{if } S = T \\ 0 & \textit{if } S \neq T \end{cases}$$

Proof.

$$\langle \chi_S, \chi_T \rangle = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} (-1)^{S\cdot x + T\cdot x} = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} (-1)^{(S\Delta T)\cdot x}$$

where Δ is the symmetric different between two sets S and T.
$S\Delta T = \emptyset$ if and only if $S = T$. Hence, our goal follows immediately from Lemma 1.3. □

1.3 Fourier Basis

Theorem 1.5. *The set of all χ_S defines an orthonormal basis for the space of all real-valued function on $\{0,1\}^n$*

Proof. From Theorem 1.4, the set of all χ_S is an orthonormal set. Also, there are 2^n different χ_S. Hence, the set of all χ_s must be an orthonormal basis for the space of all real-valued functions on $\{0,1\}^n$. □

The set of all χ_S is called the *the Fourier basis.*

1.4 Fourier Transform

Definition 1.6. For each $S \subseteq [n]$, we define the Fourier transform of f at S as following:

$$\widehat{f}(S) := \mathbb{E}[f \cdot \chi_S] = \langle f, \chi_S \rangle$$

Definition 1.7 (Fourier transform)**.** The mapping $\mathcal{F} : f \mapsto \widehat{f}$ is called the Fourier transform.

If we view functions $f, \widehat{f}$ as N-dimensional vectors, then we can write the Fourier transform as the product of f and some matrix F as following:

$$\begin{aligned}
\mathcal{F}(f) &= f \cdot F \\
&= \frac{1}{N}\left(f(0), f(1), \cdots, f(N-1)\right) \begin{pmatrix} \chi_0(0) & \chi_1(0) & \cdots & \chi_{N-1}(0) \\ \chi_0(1) & \chi_1(1) & \cdots & \chi_{N-1}(1) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_0(N-1) & \chi_1(N-1) & \cdots & \chi_{N-1}(N-1) \end{pmatrix} \\
&= (\widehat{f}(0), \widehat{f}(1), \cdots, \widehat{f}(N-1)) \\
&= \widehat{f}
\end{aligned}$$

where $F_{ij} = \frac{\chi_i(j)}{N}$. Since $\chi_i(j) = \chi_j(i)$, F is symmetric.

Lemma 1.8. F is invertible

Proof. We have

$$(F \cdot F)_{ij} = \frac{1}{N^2} \sum_{k=0}^{N-1} \chi_i(k) \cdot \chi_j(k) = \frac{1}{N} \langle \chi_i, \chi_j \rangle$$

Based on Theorem 1.4, it is easy to see that

$$(F \cdot F)_{ij} = \begin{cases} \frac{1}{N} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

So $F \cdot F = N \cdot I$, which implies that F is invertible □

Theorem 1.9. *The mapping $\mathcal{F}$ is linear.*

Proof. This follows from the properties of inner product. For any S,

$$\widehat{af + bg}(S) = \langle af + bg, \chi_S \rangle = \langle af, \chi_S \rangle + \langle bg, \chi_S \rangle = a\langle f, \chi_S \rangle + b\langle g, \chi_s \rangle = a\widehat{f}(S) + b\widehat{g}(S)$$

Thus, $\widehat{af + bg} = a\widehat{f} + b\widehat{g}$, which means $\mathcal{F}$ is linear.
Here is another way to show that.

$$\widehat{af + bg} = \mathcal{F}(af + bg) = (af + bg)F = afF + bgF = a(fF) + b(gF) = a\widehat{f} + b\widehat{g}$$

□

Theorem 1.10. *The linear map $\mathcal{F}$ is a bijection.*

Proof. It suffices to show that F is invertible, which follows immediately from Lemma 1.8. □

1.5 Parseval's Identity

Since the set of χ_S forms an orthonormal basis,

$$f = \sum_S \widehat{f}(S)\chi_S. \tag{1.1}$$

Hence,

$$\langle f, g\rangle = \sum_S \widehat{f}(S)\widehat{g}(S) \tag{1.2}$$

In particular, when f = g we get Parseval's identity:

$$\|f\|_2^2 = \sum_S \widehat{f}(S)^2 \tag{1.3}$$

This also implies:

$$\|f-g\|_2^2 = \sum_S (\widehat{f}(S) - \widehat{g}(S))^2 \tag{1.4}$$

1.6 Convolution

Definition 1.11. Given any two function f and $g : \{0,1\}^n \to \mathbb{R}$, the convolution of $f * g : \{0,1\}^n \to \mathbb{R}$ is defined as

$$(f * g)(x) := \frac{1}{2^n} \sum_{y \in \{0,1\}^n} f(x \oplus y) g(y)$$

Theorem 1.12. *If X and Y are n-bits random independent variables with probability distributions f and g, respectively, then $2^n(f * g)$ is the distribution of the random variable $Z = X \oplus Y$.*

Proof.

$$\begin{aligned} Pr[Z = z] &= Pr[X = z \oplus Y] \\ &= \sum_{y \in \{0,1\}^n} Pr[X = z \oplus y | Y = y] \\ &= \sum_{y \in \{0,1\}^n} Pr[X = z \oplus y] \cdot Pr[Y = y] \\ &= \sum_{y \in \{0,1\}^n} f((z \oplus y) \cdot g(y) \\ &= 2^n (f * g)(z) \end{aligned}$$

□

Theorem 1.13. *For every $S \subseteq [n]$,*

$$\widehat{f * g}(S) = \widehat{f}(S) \cdot \widehat{g}(S)$$

Proof.

$$\begin{aligned}\widehat{f*g}(S) &= \frac{1}{2^n}\sum_x (f*g)(x)\chi_S(x)\\ &= \frac{1}{2^n}\sum_x\left(\frac{1}{2^n}\sum_y f(x\oplus y)g(y)\right)\chi_S(x)\\ &= \frac{1}{2^{2n}}\sum_x\sum_y f(x\oplus y)g(y)\chi_S(x\oplus y)\chi_S(y)\\ &= \frac{1}{2^n}\sum_x f(x\oplus y)\chi_S(x\oplus y)\left(\frac{1}{2^n}\sum_y g(y)\chi_S(y)\right)\\ &= \widehat{f}(S)\cdot\widehat{g}(S)\end{aligned}$$

□

Intuitively, the convolution $f * g$ is the product of the Fourier transforms of f and g.

Theorem 1.14. *Let V be a subspace of dimension k of $\{0,1\}^n$ and let $V^{\perp}$ be the dual of V. Define*

$$f(x) = \begin{cases} \frac{1}{2^k} & \textit{if } x \in V \\ 0 & \textit{otherwise.} \end{cases}$$

.

Then

$$\widehat{f}(S) = \begin{cases} \frac{1}{N} & \textit{if } S \in V^{\perp} \\ 0 & \textit{otherwise.} \end{cases}$$

.

Proof. Suppose $v_1, v_2, ..., v_k$ is a basis of V.

Claim 1.15. $V = \langle v_1\rangle \oplus \langle v_2\rangle \oplus ... \oplus \langle v_k\rangle$

Claim 1.16. $\langle v_1\rangle^{\perp} \cap ... \cap \langle v_k\rangle^{\perp} = \langle v_1, ..., v_k\rangle^{\perp} = V^{\perp}$

Let $f_i = \begin{cases} \frac{1}{2} & \text{if } S \in \langle v_i\rangle \\ 0 & \text{otherwise} \end{cases}$, then $\widehat{f_i} = \begin{cases} \frac{1}{N} & \text{if } S \in \langle v_i\rangle^{\perp} \\ 0 & \text{otherwise.} \end{cases}$.

From above claims, we immediately obtain following result.

Claim 1.17. $f = f_1 \oplus f_2 \oplus ... \oplus f_k$

Hence,

$$\widehat{f}(S) = N^{k-1}\widehat{f_1}(S)...\widehat{f_k}(S)$$

If $S \in V^{\perp}$, then $S \in \langle v_i\rangle^{\perp}$ for every i, so $\widehat{f}(S) = N^{k-1}\cdot(\frac{1}{N})^k = \frac{1}{N}$.

If $S \notin V^{\perp}$, then there exits some i such that $S \notin \langle v_i\rangle^{\perp}$, which implies $\widehat{f_i}(S) = 0$. Hence, $\widehat{f}(S) = 0$ □

Chapter 2

Min Entropy

Let $X = (x_0, x_1, ..., x_{N-1})$ be a distribution function of a random variable over $\{0,1\}^n$, where $N := 2^n$.

Definition 2.1 (Min Entropy)**.** We define the min entropy of X as follow.

$$H_\infty(X) := \min_i(-\log x_i)$$

This implies that if $H_\infty(X) \geq k$ then $x_i \leq \frac{1}{2^k}$ for every $0 \leq i \leq N-1$

Theorem 2.2 (Collision Probability)**.** *If we sample X twice, then the probability we get the same result, denoted $Col(X)$, is $\sum\limits_{i=0}^{N-1} {x_i}^2 = N \cdot \|X\|_2^2$.*

Definition 2.3 (Flat Distribution)**.** A probability distribution function $f\colon \{0,1\}^n \to (0,1)$ is a *T-flat* if there $\exists\ S \subseteq \{0,1\}^n$ such that $|S| = T$ and $f(x) = \begin{cases} \frac{1}{T} & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}$

Lemma 2.4. For every $\alpha \geq \beta$, every α-flat distribution can be written as the sum of β-flat distributions.

Theorem 2.5. *For every integer $k \geq 0$, if $H_\infty(X) \geq k$, then $X = \sum \alpha_i X_i$, where each X_i is a 2^k-flat, $\alpha_i \in [0,1]$ for every i, and $\sum\limits_i \alpha_i = 1$.*

Proof. Let S be the set of all the probability distributions X with $H_\infty(X) \geq k$, then S is a compact (closed and bounded) convex set in $\mathbb{R}^N$.

Claim 2.6. The set of all 2^k-flat distributions is the set of all extreme points of S.

First, every 2^k-flat distribution X is an extreme point of S since if $X = \alpha Y + (1-\alpha)Z$ for some $Y, Z \in S$ and $\alpha \in [0,1]$, then $X = Y = Z$. Next, we want to show that for any X, which is not a 2^k-flat distribution, X is not an extreme point of S. Since X is not a 2^k-flat distribution, there exist $i < j$ such that $0 < x_i, x_j < \frac{1}{2^k}$. So we can choose $\delta > 0$ such that $x_i + \delta, x_j + \delta \leq \frac{1}{2^k}$ and $x_i - \delta, x_j - \delta \geq 0$. Now let $y_k = z_k = x_k$ for every $k \neq i, k \neq j$, $y_i = x_i + \delta$, $y_j = x_j - \delta$, $z_i = x_i - \delta$, and $z_j = x_j + \delta$. Then $Y, Z \in S$ and $X = \frac{1}{2}Y + \frac{1}{2}Z$, which implies that X is not an extreme point.

Back to the problem, since S is a compact convex set, the set of all convex combinations of its vertices is identical to S. Hence, every distribution X with $H_\infty(X) \geq k$ can be written as a convex combination of k-flat distributions.

□

Theorem 2.7. *If $H_\infty(X) \geq k$, then $Col(X) \leq \frac{1}{2^k}$.*

Proof. By theorem 2.5, we can write X as $X = \sum_i \alpha_i X_i$, where each X_i is a 2^k-flat, $\sum \alpha_i = 1$, and $\alpha_i \in [0,1]$for every i. It is obvious that $Col(X_i) = \|X_i\|_2^2 = \frac{1}{2^k}$.
Collision functions are convex, so by Jensen's inequality,

$$Col(X) = Col\left(\sum_i \alpha_i X_i\right) \leq \sum_i \alpha_i \cdot Col(X_i) = \sum_i \alpha_i \frac{1}{2^k} = \frac{1}{2^k}\sum_i \alpha_i = \frac{1}{2^k}$$

□

Theorem 2.8. *If $H_\infty(X) \geq k$, then $\sum_S \widehat{X}(S)^2 \leq \frac{1}{N \cdot 2^k}$.*

This follow immediately from the Parseval's identity.

Definition 2.9 (Small Bias Distribution)**.** Let $\mathcal{D}$ be a probability distribution function over $\{0,1\}^n$. We say that $\mathcal{D}$ is α-bias if $\widehat{D}(S) \leq \frac{\alpha}{N}$.

Definition 2.10. *Statistical Different* between two distributions A and B is defined as follow:

$$SD(A,B) = \frac{1}{2}\sum_i |a_i - b_i|$$

Theorem 2.11. *Let $\mathcal{D}$ be a small bias distribution with $\widehat{\mathcal{D}}(S) \leq \frac{\alpha}{N}$ for all S, let $\mathcal{M}$ be a min entropy source such that $H_\infty(\mathcal{M}) \geq k$, and let $\mathcal{U}$ be the uniform distribution over n-bits string. Then*

$$SD(\mathcal{D} \oplus \mathcal{M}, \mathcal{U}) \leq \frac{\alpha\sqrt{N}}{2^{1+k/2}}$$

Proof.

$$\begin{aligned}
SD(\mathcal{D} \oplus \mathcal{M}, \mathcal{U}) &= \frac{1}{2}\sum_i |(\mathcal{D} \oplus \mathcal{M})(i) - \mathcal{U}(i)| \\
&\leq \frac{1}{2}\sqrt{N\sum_i [(\mathcal{D} \oplus \mathcal{M})(i) - \mathcal{U}(i)]^2} \\
&= \frac{1}{2}\sqrt{N^2 \cdot \|(\mathcal{D} \oplus \mathcal{M}) - \mathcal{U}\|_2^2} \\
&= \frac{N}{2}\sqrt{\sum_S [\widehat{\mathcal{D} \oplus \mathcal{M}}(S) - \mathcal{U}(S)]^2} \\
&= \frac{N}{2}\sqrt{\sum_{S \neq \emptyset} \widehat{\mathcal{D} \oplus \mathcal{M}}(S)^2}
\end{aligned}$$

By convolution,

$$\begin{aligned}\sum_{S\neq\emptyset}\widehat{\mathcal{D}\oplus\mathcal{M}}(S)^2 &= \sum_{S\neq\emptyset} N^2\cdot\widehat{\mathcal{D}*\mathcal{M}}(S)^2\\ &= N^2\sum_{S\neq\emptyset}\widehat{\mathcal{D}}(S)^2\cdot\widehat{\mathcal{M}}(S)^2\\ &\leq N^2\cdot\sum_{S\neq\emptyset}(\frac{\alpha}{N})^2\cdot\widehat{\mathcal{M}}(S)^2\\ &= \alpha^2\cdot\sum_{S\neq\emptyset}\widehat{\mathcal{M}}(S)^2\\ &\leq \frac{\alpha^2}{N\cdot 2^k}\end{aligned}$$

Hence,

$$SD(\mathcal{D}\oplus\mathcal{M},\mathcal{U})\leq\frac{\alpha\sqrt{N}}{2^{1+k/2}}$$

□

Theorem 2.12. *Let M be a distribution with min entropy k over $\{0,1\}^n$, let $G_0\sim m\times 1$, $G\sim m\times n$, and let X be a uniform distribution over $\{0,1\}^m$. Then*

$$2SD\{(XG_0, XG\oplus M, G_0, G), (U, XG\oplus M, G_0, G)\}\leq\sqrt{\frac{2N}{MK}}$$

Proof. For convenience, let $A_{G_0,G} = (XG_0, XG\oplus M|G_0,G)$, $A'_{G_0,G} = (XG_0, XG|G_0,G)$ and $B_{G_0,G} = (U, XG\oplus M|G_0,G)$.

Claim 2.13. For any $S\subseteq[n+1]$, $\widehat{A_{G_0,G}}(S) = \widehat{B_{G_0,G}}(S)$ if $S_1 = 0$, and $\widehat{B_{G_0,G}}(S) = 0$ if $S_1 = 1$.

Claim 2.14. For any set $S\subseteq[n+1]$ with $S_1 = 1$,

$$\mathop{\mathbb{E}}_{G_0,G}[\widehat{A'_{G_0,G}}(S)^2]\leq\frac{1}{(2N)^2\cdot M}$$

Proof.

$$\widehat{A'_{G_0,G}}(S) = \mathbb{E}[A'_{G_0,G}(x)\cdot\chi_S(x)] = \frac{1}{2N}\text{bias}_S(A'_{G_0,G})$$

$$\begin{aligned}\text{where bias}_S(A'_{G_0,G}) &= \left|A'_{G0,G}\{x:\bigoplus_{i\in S}xG_{i-1}=0\} - A'_{G0,G}\{x:\bigoplus_{i\in S}xG_{i-1}=1\}\right|\\ &= \left|A'_{G0,G}\{x:x\cdot(\bigoplus_{i\in S}G_{i-1})=0\} - A'_{G0,G}\{x:x\cdot(\bigoplus_{i\in S}G_{i-1})=1\}\right|\end{aligned}$$

If $G_0 = \underset{i\in S_{>1}}{\oplus} G_{i-1}$, then $\text{bias}_S(A'_{G_0,G}) = 1$, and if $G_0 \neq \underset{i\in S_{>1}}{\oplus} G_{i-1}$ then $\text{bias}_S(A'_{G_0,G}) = 0$.

$$Pr[G_0 = \underset{i\in S_{>1}}{\oplus}G_{i-1}] \le Pr[G_0 \in \langle G_{i-1} : i \in S\rangle] \le \frac{1}{M}$$

$$Pr[G_0 \neq \underset{i\in S_{>1}}{\oplus}G_{i-1}] \ge Pr[G_0 \in \langle G_{i-1} : i \in S\rangle] \ge 1 - \frac{1}{M}$$

$$\underset{G_0,G}{\mathbb{E}}[\widehat{A'_{G_0,G}}(S)^2] \le \frac{1}{M}\cdot\left(\frac{1}{2N}\right)^2 = \frac{1}{(2N)^2\cdot M} \qquad \square$$

$$\begin{aligned}
&2SD\{(XG_0, XG\oplus M, G_0, G), (U, XG\oplus M, G_0, G)\}\\
=&\underset{G_0,G}{\mathbb{E}}[2SD(A_{G_0,G}, B_{G_0,G})]\\
=&\underset{G_0,G}{\mathbb{E}}\left[\sum_{i,j}|A_{G_0,G}(i) - B_{G_0,G}(i)|\right]\\
\le&\underset{G_0,G}{\mathbb{E}}\left[\sqrt{2N\cdot 2N\cdot \|A_{G_0,G} - B_{G_0,G}\|_2^2}\right] && \text{[by Cauchy Schwartz]}\\
\le&2N\underset{G_0,G}{\mathbb{E}}\left[\sqrt{\sum_S(\widehat{A_{G_0,G}}(S) - \widehat{B_{G_0,G}}(S))^2}\right] && \text{[by Parseval's Idenity]}\\
\le&2N\cdot\sqrt{\underset{G_0,G}{\mathbb{E}}\left[\sum_S(\widehat{A_{G_0,G}}(S) - \widehat{B_{G_0,G}}(S))^2\right]} && \text{[by Jensen's Inequality]}\\
=&2N\cdot\sqrt{\sum_S\underset{G_0,G}{\mathbb{E}}\left[(\widehat{A_{G_0,G}}(S) - \widehat{B_{G_0,G}}(S))^2\right]} && \text{[by linearity of expectation]}\\
=&2N\cdot\sqrt{\sum_{S_1=1}\underset{G_0,G}{\mathbb{E}}\left[\widehat{A_{G_0,G}}(S)^2\right]} && \text{[by Claim 2.13]}\\
=&2N\cdot\sqrt{\sum_{S_1=1}\underset{G_0,G}{\mathbb{E}}\left[(\widehat{(0,M)}(S)\oplus(XG_0,\widehat{XG}|G_0,G)(S))^2\right]}\\
=&2N\cdot\sqrt{\sum_{S_1=1}\underset{G_0,G}{\mathbb{E}}\left[(2N)^2\cdot\widehat{(0,M)}(S)^2\cdot(XG_0,\widehat{XG}|G_0,G)(S)^2\right]} && \text{[by convolution]}\\
=&4N^2\cdot\sqrt{\sum_{S_1=1}\widehat{(0,M)}(S)^2\cdot\underset{G_0,G}{\mathbb{E}}\left[\widehat{A'_{G_0,G}}(S)^2\right]}\\
\le&4N^2\cdot\sqrt{\sum_{S_1=1}\widehat{(0,M)}(S)^2\cdot\frac{1}{(2N)^2\cdot M}} && \text{[by Claim 2.14]}\\
\le&4N^2\cdot\sqrt{\frac{1}{2N\cdot K}\cdot\frac{1}{(2N)^2\cdot M}} = \sqrt{\frac{2N}{MK}} && \square
\end{aligned}$$

□

Chapter 3

Generalized Fourier Transform

The entire treatment for Fourier analysis in chapter 1 can be generalized to functions from any finite abelian group to the set of complex numbers.

3.1 Notation

$\mathbb{C}$: the set of complex numbers.
$\mathbb{G}$: a finite abelian group of order n.
$\mathbb{C}^* = \mathbb{C} \setminus \{0\}$: multiplicative group of complex numbers.
$\mathbb{F}_q$: the field of size q
For a complex number x, we use $\bar{x}$ to denote the complex conjugate of x.

3.2 Inner Product

Let $f : G \to \mathbb{C}, \text{and } g : G \to \mathbb{C}$ be two functions from a finite abelian group to the complex numbers.

Definition 3.1. The inner product of f and g is defined as

$$\langle f, g \rangle = \mathop{\mathbb{E}}_{x}[f(x)\overline{g(x)}]$$

Definition 3.2. The L^p norm of f is defined as $\|f\|_p = (\mathbb{E}_x |f(x)|^p)^{1/p}$

Definition 3.3. The ℓ^p norm of f is defined as $\|f\|_{\ell^p} = (\sum\limits_{x \in G} |f(x)|^p)^{1/p}$

Definition 3.4. The ℓ^∞ norm of f is defined as $\|f\|_{\ell^\infty} = \max_x |f(x)|$

Some basic relations between these norms.

Proposition 3.5 (Triangle Inequality). $|\langle l, g \rangle| \leq \|f\|_1 \|g\|_{\ell^\infty}$

Proposition 3.6 (Cauchy Schwartz Inequality). For any two functions $f, g : G \to \mathbb{C}$,

$$|\langle f, g \rangle| \leq \|f\|_2 \cdot \|g\|_2$$

3.3 Characters

Definition 3.7. A character of G is a homomorphism $\chi : G \to \mathbb{C}^*$.

χ is a trivial (principal) character if $\chi(a) = 1$ for every $a \in G$.
Clearly,

$$\chi(a+b) = \chi(a)\chi(b) \text{ for every } a, b \in G$$

Hence,

$$\chi(a)^n = \chi(na) = \chi(0) = 1$$

Intuitively, the values of χ are n^{th} roots of unity. So,

$$\overline{\chi(a)} = \chi(a)^{-1} = \chi(-a)$$

Proposition 3.8. For any non-trivial character χ,

$$\sum_{a \in G} \chi(a) = 0$$

Proof. Since χ is a non-trivial character, there exists $b \in G$ such that $\chi(b) \neq 1$. We have

$$\chi(b) \cdot \sum_{a \in G} \chi(a) = \sum_{a \in G} \chi(a)\chi(b) = \sum_{a \in G} \chi(a+b) = \sum_{a \in G} \chi(a)$$

Hence, $\sum_{a \in G} \chi(a) = 0$ □

Lemma 3.9. For any two characters χ and ψ of G,

$$\langle \chi, \psi \rangle = \begin{cases} 1 & \text{if } \chi = \psi \\ 0 & \text{otherwise.} \end{cases}$$

Proof. If $\chi = \psi$, $\langle \chi, \chi \rangle = (1/|G|) \sum_{x \in G} \chi(x)\overline{\chi(x)} = (1/|G|) \sum_{x \in G} 1 = 1$.
If $\chi \neq \psi$, let $(\chi\overline{\psi})(x) = \chi(x)\overline{\psi(x)}$, then $(\chi\overline{\psi})$ is a homomorphism since

$$(\chi\overline{\psi})(x+y) = \chi(x+y)\overline{\psi(x+y)} = \chi(x)\chi(y)\overline{\psi(x)\psi(y)} = \chi(x)\overline{\psi(x)}\chi(y)\overline{\psi(y)} = (\chi\overline{\psi})(x)(\chi\overline{\psi})(y)$$

Thus, by Lemma 3.8

$$\langle \chi, \psi \rangle = (1/|G|) \sum_{x \in G} \chi(x)\overline{\psi(x)} = (1/|G|) \sum_{x \in G} (\chi\overline{\psi})(x) = 0$$

. □

3.4 Bi-linear Maps

Definition 3.10. We say that a map $e : G \times G \to \mathbb{C}$ is a bilinear map if it is a homomorphism in each variable. It is a non-degenerate if for every x, both $e(\cdot, x)$ and $e(x, \cdot)$ are non-trivial. It is symmetric if $e(x, y) = e(y, x)$ for every x, y.

Lemma 3.11. Let e be the map that maps $(x, y) \mapsto \exp(2\pi xyi/r)$, then e is a symmetric non-degenerate bilinear map.

Definition 3.12. Let $G^\wedge$ denote the set of all characters of G.

Lemma 3.13. $G^\wedge$ is an abelian group under pointwise multiplication.

Lemma 3.14. For every abelian group G, there exists a symmetric non-degenerate bilinear $e : G \times G \to \mathcal{C}$

Now fix any symmetric, non-degenerate, bilinear map e. For every $x \in G$, let e_x denote the character $e(x, \cdot)$. The map $x \mapsto e_x$ can be shown to be an isomorphism from G to $G^\wedge$. Thus, we have following lemma.

Lemma 3.15. Every finite abelian group G is isomorphic to its character group $G^\wedge$

Theorem 3.16. *$G^\wedge$ forms an orthonormal basis in $\mathbb{C}^n$*

This follows immediately from Lemmas 3.4 and 3.9.

3.5 Fourier Transform

Definition 3.17. We define $\widehat{f} : G \to \mathbb{C}$ as

$$\widehat{f}(x) := \langle f, e_x \rangle$$

$\widehat{f}$ is called the Fourier transform of f. From Theorem 3.10, every function $f \in \mathbb{C}^n$ can be written as linear combinations of characters.

$$f(x) = \sum_{y \in G} \widehat{f}(y) \cdot e_y(x)$$

Definition 3.18. The mapping $\mathcal{F} : f \mapsto \widehat{f}$ is called the Fourier transform.

Fact 3.19. $\mathcal{F}$ is a linear, bijective mapping.

Proposition 3.20 (Preservation of Inner Product). $\langle f, g \rangle = |G| \langle \widehat{f}, \widehat{g} \rangle$

Proof.

$$\begin{aligned}
\langle f, g\rangle &= \frac{1}{|G|}\sum_{x\in G} f(x)\overline{g(x)} \\
&= \frac{1}{|G|}\sum_{x\in G}\left(\sum_{y\in G}\widehat{f}(y)e_y(x)\right)\left(\sum_{z\in G}\overline{\widehat{g}(z)e_z(x)}\right) \\
&= \frac{1}{|G|}\sum_{x\in G}\sum_{y\in G}\sum_{z\in G}\widehat{f}(y)\overline{\widehat{g}(z)}e_y(x)\overline{e_z(x)} \\
&= \frac{1}{|G|}\sum_{y\in G}\sum_{z\in G}\widehat{f}(y)\overline{\widehat{g}(z)}\sum_{x\in G}e_y(x)\overline{e_z(x)} \\
&= \sum_{y\in G}\sum_{z\in G}\widehat{f}(y)\overline{\widehat{g}(z)}\langle e_y, e_z\rangle \\
&= \sum_{y\in G}\widehat{f}(y)\overline{\widehat{g}(y)} \\
&= |G|\langle\widehat{f}, \widehat{g}\rangle
\end{aligned}$$

□

If we let $f = g$ in the above equation, we will get Parseval's identity as follows.

Proposition 3.21 (Paserval's Identity)**.** $\|f\|_{L^2} = \|\widehat{f}\|_{\ell^2}$

This follows from Parseval's identity and basic inequalities between norms.

Proposition 3.22. $\|f\|_{\ell^1} \leq \|\widehat{f}\|_{\ell^\infty}$

Proposition 3.23 (Fourier Inversion)**.** $f(x) = |G|\hat{\hat{f}}(-x)$

Proof.

$$\begin{aligned}
|G|\hat{\hat{f}}(-x) &= |G|\langle\widehat{f}, e_{-x}\rangle \\
&= \sum_{y\in G}\widehat{f}(y)e_x(y) \\
&= \frac{1}{|G|}\sum_{y\in G}\langle f, e_y\rangle e_x(y) \\
&= \frac{1}{|G|}\sum_{y\in G}\sum_{z\in G} f(z)\overline{e_z(y)}e_x(y) \\
&= \sum_{z\in G} f(z)\langle e_x, e_z\rangle \\
&= f(x)
\end{aligned}$$

□

Chapter 4

XOR Lemma over Finite Abelian Groups

Here are some facts needed to prove XOR lemma.

Fact 4.1. If $\sigma : (G, +) \to (H, *)$ is an onto homomorphism, then for all non-trivial character ϕ of H, $\phi \circ \sigma$ is a non-trivial character of G.

$$(\phi \circ \sigma)(x+y) = \phi(\sigma(x+y)) = \phi(\sigma(x) * \sigma(y)) = \phi(\sigma(x)) \cdot \phi(\sigma(y)) = (\phi \circ \sigma)(x) \cdot (\phi \circ \sigma)(y)$$

Fact 4.2. For every character ϕ of H, and for every $\sigma : G \to H$,

$$\text{bias}_\phi(\sigma(X)) = |H| \cdot |\langle \phi, \sigma(X) \rangle| = |G| \cdot |\langle \phi \circ \sigma, X \rangle|$$

This follows from the fact that

$$\sigma(X)(y) = \sum_{x:\sigma(x)=y} X(x)$$

Fact 4.3. If e_x is a character of G, then $\|\widehat{e_x}\|_{L^1} = 1/|G|$.

$$\|\widehat{e_x}\|_{L^1} = \mathbb{E}_y|\widehat{e_x}| = \mathbb{E}_y|\langle e_x, e_y \rangle| = 1/|G|(|\langle e_x, e_x \rangle| + \sum_{y \neq x} |\langle e_x, e_y \rangle|) = 1/|G|$$

Fact 4.4. $\|\sigma(X) - U\|_{\ell^1} \leq \|\sigma(X) - \sigma(U)\|_{\ell^1} + \|\sigma(U) - U\|_{\ell^1}$

Following result is a generalization of Vazirani's XOR lemma.

Lemma 4.5 (XOR Lemma)**.** Let X be a distribution on a finite abelian group G such that $\text{bias}_\psi(X) = |\mathbb{E}[\psi(X)]| \leq \epsilon$ for every nontrivial character ψ. Then X is $\epsilon\sqrt{|G|}$-close to the uniform distribution: $\|X - U\|_{\ell^1} \leq \epsilon\sqrt{|G|}$

Proof. The assumption on X is equivalent to $\|X - U\|_{\ell^\infty} \leq \epsilon/\sqrt{|G|}$. Then, we need to apply proposition ... to get the result. □

Note that the upper bound we have just get is propositional to the square root of the size of G. If ϵ is not small enough, then this bound is not good. In fact, we can't get a better bound for the same number of bits, which is $|G|$, since our inequality is tight,i.e, we can construct a distribution X so that the equality happens. Thus, if we want to have some better bound, we have to trade-off the number of bits. Formally, we have the following lemma.

Lemma 4.6 (XOR Lemma for Cyclic Groups)**.** For every cyclic group $G = \mathbb{Z}_N$ and for every integer $M \leq N$, there exists an efficiently computable function $\sigma : \mathbb{Z}_N \to \mathbb{Z}_M = H$ with following property: If X is a random variable over $\mathbb{Z}_N$ such that for every non-trivial character $\psi : \mathbb{Z}_M \to \mathbb{C}^*$, we have $\mathrm{bias}_\psi(X) \leq \epsilon$, then

$$\|\sigma(X) - U\|_{\ell^1} \leq O(\epsilon\sqrt{M}\log N) + O(M/N)$$

Some observations: Based on the Fact 4.1, if there exists an onto homomorphism $\sigma : G \to H$, then we just need to apply Lemma 4.5 to get the result since biases of X give bounds on the biases of $\sigma(X)$ by the Fact 4.2. However, such a homomorphism doesn't exist for every M. For example, if $G = \mathbb{Z}_p$ and $M > p/2$, G has no non-trivial subgroup. Instead, we can find a $\sigma = x \mod M$ that approximates a homomorphism. By the Fact 4.4, if we can show that $\sigma(U)$ is close to uniform and that $\sigma(X)$ is close to $\sigma(U)$, then we are done.

Lemma 4.7. Let U be the uniform distribution over $\mathbb{Z}_M$, and let $\sigma : \mathbb{Z}_N \to \mathbb{Z}_M$ such that $\sigma(x) = x \mod M$. Then,

$$\|\sigma(U) - U\|_{\ell^1} \leq 2M/N$$

Proof. Suppose $N = qM + r$ where q, r are the quotient and remainder respectively. Then,

$$\Pr[\sigma(U) = i | i < r] = (q+1)/N, \text{and } \Pr[\sigma(U) = i | i >= r] = q/N$$

Thus,

$$\begin{aligned}
|\sigma(U) - U\|_{\ell^1} &= r\left|\frac{(q+1)}{N} - \frac{1}{M}\right| + (M - r)\left|\frac{q}{N} - \frac{1}{M}\right| \\
&= \frac{2r(M-r)}{MN} \\
&\leq 2M/N
\end{aligned}$$

□

To bound $\|\sigma(X) - \sigma(U)\|_{\ell^1}$, we just need to bound $\|\widehat{\sigma(X) - \sigma}(U)\|_{\ell^\infty}$, which is the maximum Fourier coefficient of $\sigma(X) - \sigma(U)$. Notice that, for any non-trivial character ψ,

$$\begin{aligned}
|\langle\psi, \sigma(X) - \sigma(U)\rangle| &= \frac{N}{M}\langle\psi \circ \sigma, X - U\rangle| && \text{[by Fact 4.2]} \\
&= \frac{N^2}{M}|\langle\widehat{\psi \circ \sigma}, \widehat{X - U}\rangle| \\
&\leq \frac{N^2}{M}\|\widehat{\psi \circ \sigma}\|_{L^1}\|\widehat{X - U}\|_{\ell^\infty} && \text{[by triangle inequality]}
\end{aligned}$$

And $\langle 1, \sigma(X) - \sigma(U)\rangle| = 0$. Hence, if $\|\psi \circ \sigma\|_{L^1}$ is small, so is $\|\widehat{\sigma(X) - \sigma(U)}\|_{\ell^\infty}$. The following lemma will show that $\|\psi \circ \sigma\|_{L^1}$ is small if $N > M$

Lemma 4.8. $\|\widehat{\psi \circ \sigma}\|_{L^1} \leq O(\log N)/N$

Proof. Let $\rho : G \to H$ such that $\rho(x) = \exp(2\pi i x)$, then ρ is a homomorphism. Since ϕ is a character of $H = \mathbb{Z}_M$, for every $y \in H$, $\phi(y) = \rho(wy/M)$ for some $w \in H$. Clearly, for every $x \in G$, $\phi(\sigma(x)) = \phi(x \mod M) = \rho(wx/M)$. Let $wN/M = a = c + d$, where c is an integer and $d \in [0, 1]$.

$$\begin{aligned}
\|\widehat{\psi \circ \sigma}\|_{L^1} &= 1/N \sum_{t \in \mathbb{Z}_N} |\widehat{\phi \circ \sigma}(t)| \\
&= 1/N \sum_{t \in \mathbb{Z}_N} |\langle e_t, \phi \circ \sigma\rangle| \\
&= 1/N^2 \sum_{t \in \mathbb{Z}_N} \left| \sum_{x \in \mathbb{Z}_N} e_t(x) \cdot \overline{\phi(\sigma(x))} \right| \\
&= 1/N^2 \sum_{t \in \mathbb{Z}_N} \left| \sum_{x \in \mathbb{Z}_N} \rho(tx/N) \cdot \rho(-wx/M) \right| \\
&= 1/N^2 \sum_{t \in \mathbb{Z}_N} \left| \sum_{x \in \mathbb{Z}_N} \rho\left(\frac{x(tM - wN)}{MN}\right) \right| \\
&\leq 1/N + 1/N^2 \sum_{t \in \mathbb{Z}_N, t \neq a} \left| \left(\rho\left(\frac{N(tM - wN)}{MN}\right) - 1\right) \Big/ \left(\rho\left(\frac{tM - wN}{MN}\right) - 1\right) \right| \\
&\leq 1/N + 1/N^2 \sum_{t \in \mathbb{Z}_N, t \neq a} \left| 2 \Big/ \left(\rho\left(\frac{t - a}{N}\right) - 1\right) \right| \\
&= 1/N + 1/N^2 \sum_{t \in \mathbb{Z}_N, t \neq d} \left| 2 \Big/ \left(\rho\left(\frac{t - d}{N}\right) - 1\right) \right|
\end{aligned}$$

Notice that $\left|\rho\left(\frac{t-d}{N}\right) - 1\right|$ is the distance between two points $(1, 0)$ and $(\cos(2\pi(t-d)/N), \sin(2\pi(t-d)/N))$ in the unit circle, and that $\left|\rho\left(\frac{t-d}{N}\right) - 1\right| \geq |\sin(2\pi(t-d)/N))| \geq |2\pi(t-d)/N)|$ if $|2\pi(t-d)/N)| \leq \pi/2$. Thus, let choose $0 < r < 1/4$, then

- if $t \in [0, rN)$ then $|2\pi(t-d)/N)| \leq \pi/2$,
- if $t \in ((1-r)N, rN]$ then $\left|\rho\left(\frac{t-d}{N}\right) - 1\right| = \left|\rho\left(\frac{-t'-d}{N}\right) - 1\right| \geq 2\pi(t' + d)/N$ since $(t' + d)/N < 1/4$, where $t' = N - t \in [0, rN)$
- if $t \in [rN, (1-r)N]$ then $\left|\rho\left(\frac{t-d}{N}\right) - 1\right| \geq \Omega(1)$

Therefore,

$$\|\widehat{\psi \circ \sigma}\|_{L^1} \leq 1/N + (1/N^2)(O(N \log N) + O(N)) + 1/N \leq O(\log N)/N$$

□

Now, we have

$$\|\sigma(\widehat{X) - \sigma}(U)\|_{\ell^\infty} \leq (N^2/M) \cdot O(\log N)/N \cdot (\epsilon/N) = O(\epsilon \log N)/M$$

Applying Lemma ...,

$$\|\sigma(X) - \sigma(U)\|_{\ell^1} \leq O(\epsilon\sqrt{M}\log N)$$

By Fact 4.4,

$$\|\sigma(X) - U\|_{\ell^1} \leq O(\epsilon\sqrt{M}\log N) + O(M/N)$$

Chapter 5

Bourgain's Extractor

5.1 Introduction and Motivation

Randomness extraction is the problem of obtaining nearly uniform bits from sources that are only weakly random. Most of the applications in cryptography, algorithms, and so on require truly random, uncorrelated bits but most easily obtainable sources of randomness in natural do not satisfy these conditions.

An extractor Ext $: \{0,1\}^n \to \{0,1\}^m$ is a function that takes input from a weak source with sufficient min entropy and produces nearly uniform bits. Unfortunately, deterministic extractor from one source is impossible. To overcome this difficulty, 2-source extractors, each with sufficient high min entropy, have been used. By probabilistic argument, there exist 2-source extractors in which each has min entropy polynomial of $\log n$. However, it is important to construct such extractors explicitly. Hadamard extractor shows that the dot product of two sources with min entropy greater than $n/2$ is a 2-source extractor. No progress was made for a long time until Bourgain broke the half barrier, constructed a 2-source extractor with min entropy slightly less than $n/2$. For the rest of this chapter, we will focus on how to construct Bourgain's extractor.

5.2 Bourgain's Extractor

5.2.1 The Hadamard Extractor

In this section, we will show that the dot product of two sources with min entropy greater than $n/2$ is a 2-source extractor.

Let $Had : \mathbb{F}^l \times \mathbb{F}^l \to \mathbb{F}$ be the dot product function, $Had(x, y) = x \cdot y$.

Theorem 5.1. *For every constant $\delta > 0$, there exists a polynomial time algorithm Had : $(\{0,1\}^n)^2 \to \{0,1\}^m$ such that if X, Y are independent $(n, (1/2+\delta)n)$ sources, then $Had(X, Y)$*

is ϵ-close to uniform distribution, i.e, $\mathbb{E}_Y[\|Had(X,Y) - U_m\|_{\ell^1}] < \epsilon$, *where* $m = \Omega(n)$ *and* $\epsilon = 2^{-\Omega(n)}$.

Proof. For any non-trivial character ψ, we want to bound

$$bias_\psi(Had(X,Y)) = |\mathbb{E}_Y[\psi(Had(X,Y))]| = \sum_{y \in \mathbb{F}^l} Y(y) \sum_{x \in \mathbb{F}^l} X(x)\psi(x \cdot y)$$

so that we can apply XOR-lemma.
Notice that

$$\begin{aligned}
\sum_{y \in \mathbb{F}^l} Y(y) \sum_{x \in \mathbb{F}^l} X(x)\psi(x \cdot y) &= \sum_{y \in \mathbb{F}^l} Y(y) \sum_{x \in \mathbb{F}^l} X(x)e(x,y) \\
&= \sum_{y \in \mathbb{F}^l} Y(y)|\mathbb{F}|^l \langle e_y, X\rangle \\
&= |\mathbb{F}|^l \sum_{y \in \mathbb{F}^l} Y(y)\overline{\widehat{X}(y)} \\
&= |\mathbb{F}|^{2l} |\langle Y, \widehat{X}\rangle|
\end{aligned}$$

Hence,

$$\begin{aligned}
bias_\psi(X,Y)^2 &\le |\mathbb{F}|^{4l} \cdot \|Y\|_2^2 \cdot \|\widehat{X}\|_2^2 && \text{[by Cauchy-Schwartz]} \\
&= |\mathbb{F}|^{3l} \cdot \|Y\|_2^2 \cdot \|X\|_2^2 && \text{[by Parseval's Identity]} \\
&\le |\mathbb{F}|^{3l} \cdot \frac{1}{|\mathbb{F}|2^{k_1}} \cdot \frac{1}{|\mathbb{F}|2^{k_2}} \\
&= 2^{n-k_1-k_2}
\end{aligned}$$

where k_1, k_2 are min-entropy of X, Y respectively. If $k_1 = k_2 = (1/2+\delta)n$, then $bias_\psi(X,Y) \le 2^{-\delta n}$. Applying XOR Lemma, we get $\|Had(X,Y) - U_m\|_{\ell^1} \le 2^{-\delta n + m/2}$ □

5.2.2 Hadamard succeeds when the sources grow with addition

Lemma 5.2. $bias_\psi(X,Y)^2 \le bias_\psi(X - X, Y)$

Proof.

$$\begin{aligned}
bias_\psi(X,Y)^2 &= \left(\sum_{y\in\mathbb{F}^l} Y(y) \sum_{x\in\mathbb{F}^l} X(x)\psi(x\cdot y)\right)^2 \\
&\leq \left(\sum_{y\in\mathbb{F}^l} Y(y)\right) \cdot \sum_{y\in\mathbb{F}^l} Y(y) \left(\sum_{x\in\mathbb{F}^l} X(x)\psi(x\cdot y)\right)^2 && \text{[by Cauchy Schwartz]} \\
&= \left|\sum_{y\in\mathbb{F}^l} Y(y) \sum_{x_1,x_2\in\mathbb{F}^l} X(x_1)X(x_2)\psi(x_1\cdot y)\psi(-x_2\cdot y)\right| && \text{[since } e(-x,y) = -e(x,y)\text{]} \\
&= \left|\sum_{y\in\mathbb{F}^l} Y(y) \sum_{x\in\mathbb{F}^l} (X-X)(x)\psi(x\cdot y)\right| \\
&= bias_\psi(X-X,Y)
\end{aligned}$$

□

Appendix A

A.1 Dual of a Vector Space

Definition A.1 (Dual space)**.** Let V be a subspace of $\{0,1\}^n$. We define the dual of V as $V^\perp = \{x \in \{0,1\}^n | x \cdot v = 0 \ \forall v \in V\}$.

Theorem A.2. *$V^\perp$ is a subspace of $\{0,1\}^n$.*

Proof. For any $x, y \in V^\perp, a \in \{0,1\}, (a \cdot x + y) \cdot v = a \cdot (x \cdot v) + y \cdot v = 0 + 0 = 0.$ □

Lemma A.3. $\sum\limits_{i:\text{even}}^{t} \binom{n}{i} = \sum\limits_{i:\text{odd}}^{t} \binom{n}{i} = 2^{t-1}$.

Theorem A.4. *For any subspace V of dimension k of $\{0,1\}^n$, there exists a unique dual space $V^\perp$ of dimension $(n-k)$.*

Proof. We will show that $|V^\perp| = 2^{n-k}$ by induction on k.
If $k = 0$, then $V = \{\mathbf{0}\}$. Clearly, $V^\perp = \{0,1\}^n$.
If $k = 1$, let $V = \{\vec{0}, v\}$. Suppose the number of $v_i = 1$ is t, then the number of x such that $x \cdot v = 0$ is $\sum\limits_{i:2|t-i} \binom{n}{i} 2^{n-t} = 2^{t-1} \cdot 2^{n-t} = 2^{n-1}$ by Lemma A.3.
Suppose that there exists a unique orthogonal subspace $V^\perp$ of dimension $(n-k+1)$ for any subspace V of dimension $k-1$ of $\{0,1\}^n$, where $k \geq 2$.
Let $V = \langle v_1, v_2, ..., v_k \rangle$, $S_1 = \langle v_1, v_2, ..., v_{k-1} \rangle$, and $S_2 = \langle v_k \rangle$. Then, $V^\perp = S_1^\perp \cap S_2^\perp$.
Suppose $dim(V^\perp) = t$. We want to show $t = n - k$.
By induction hypothesis, $dim(S_1^\perp) = n - k + 1$ and $dim(S_2^\perp) = n - 1$.
If $t \leq n-k-1$, then we need $[(n-k+1)-t]$ independent vectors to cover $S_1^\perp$ from extending $V^\perp$, and we need $[(n-1)-t]$ independent vectors to cover $S_2^\perp$ from extending $V^\perp$. Since $S_1^\perp \cup S_2^\perp \subseteq \{0,1\}^n$, we must have $[(n-k+1)-t] + [(n-1)-t] + t \leq n$, which is equivalent to $t \geq n-k$, contradiction.
If $t \geq n-k+1$, then $S_1^\perp \subseteq S_2^\perp$, this is impossible since v_k is independent from $v_1, v_2, ..., v_{k-1}$.
Thus, $t = n-k$. So $|V^\perp| = 2^{n-k}$. □

A.2 Statistical Distance between Two Joint Distributions

Theorem A.5. *Let A, B be some probability distributions over the same sample space, and let C be a probability distribution. Then*

$$SD\{(A,C),(B,C)\} = \mathop{\mathbb{E}}_{c\sim C}[SD\{(A|C=c),(B|C=c)\}]$$

Proof.

$$\begin{aligned}
SD\{(A,C),(B,C)\} &= \frac{1}{2}\sum_{i,c}|(A,C)(i,c)-(B,C)(i,c)| \\
&= \frac{1}{2}\sum_{i,c}|Pr(C=c)\cdot Pr(A=i|C=c)-Pr(C=c)\cdot Pr(B=i|C=c)| \\
&= \sum_{c\sim C}\left(Pr(C=c)\cdot\frac{1}{2}\sum_{i}|Pr(A=i|C=c)-Pr(B=i|C=c)|\right) \\
&= \sum_{c\sim C}Pr(C=c)\cdot SD\{(A|C=c),(B|C=c)\} \\
&= \mathop{\mathbb{E}}_{c\sim C}[SD\{(A|C=c),(B|C=c)\}]
\end{aligned}$$

□

Claim A.6. For any distributions C, D, for any $S = S_C S_D \subseteq [n+1]$, if $S_C = 0$, then

$$\widehat{(C,D)}(S) = \mathop{\mathbb{E}}_{c}[(\widehat{D|C=c})(S_D)]$$

Proof.

$$\begin{aligned}
\widehat{(C,D)}(S) &= \langle (C,D), \chi_S\rangle \\
&= \frac{1}{2N}\sum_{(c,d)}(C,D)(c,d)\cdot\chi_S(c,d) \\
&= \frac{1}{2N}\sum_{(c,d)}C(c)\cdot(D|C=c)(d)\cdot\chi_{S_C}(c)\cdot\chi_{S_D}(d) \\
&= \frac{1}{2N}\sum_{c}\left[C(c)\cdot\chi_{S_C}(c)\cdot\sum_{d}(D|C=c)(d)\chi_{S_D}(d)\right] \\
&= \frac{1}{2}\sum_{c}\left[C(c)\cdot\chi_{S_C}(c)\cdot(\widehat{D|C=c})(S_D)\right] \\
&= \frac{1}{2}\sum_{c}\left[C(c)\cdot(\widehat{D|C=c})(S_D)\right] \\
&= \mathop{\mathbb{E}}_{c}[(\widehat{D|C=c})(S_D)]
\end{aligned}$$

□

A.3 Group Basics

A.3.1 Notation

We reverse the variable p to denote primes.
$\mathbb{F}_p$ denotes the field of size p.
G denotes a finite abelian group.
$\mathbb{C}$ denotes the set of complex numbers.

Definition A.7. We say $\psi : G \to \mathbb{C}$ is a character if ψ is a homomorphism.

Definition A.8. We say a map $e : G \times G \to \mathbb{C}$ is a bilinear map if it is a homomorphism in each variable.

Theorem A.9. *For every abelian group G, there exists a symmetric non-degenerate bilinear $e : G \times G \to \mathcal{C}$*

A.3.2 Dual of a finite Abelian Group

Let $G^{\wedge}$ denote the set of all characters of G.

Lemma A.10. $G^{\wedge}$ is an abelian group under point-wise product operation

Theorem A.11. *Every finite abelian group G is isomorphic to its character group $G^{\wedge}$*

A.4 Product Graph

Definition A.12. $D = (a, b, c, d)$ is a 2×2 distribution graph if and only if

1. $a + b + c + d = 1$,

2. $a, b, c, d \in [0, 1]$.

Definition A.13. $G = (x, y, z, t)$ is a 2×2 product graph if and only if

1. G is a distribution graph,

2. $xt = yz$, or $x = t = 0$, or $y = z = 0$.

Let G be the space of all 2×2 product graphs and let $\mathbb{D}$ be the space of all 2×2 distribution graphs. We want to find

$$D^* = \underset{D \in \mathbb{D}}{argmax}\ dist(G, G)$$

$$m = \max_{D \in \mathbb{D}} dist(G, G)$$

Let $D = (a, b, c, d)$ be any 2×2 distribution graph. Without lost of generality, assume $a \geq d$, $b \geq c$, and $ad \geq bc$

Claim A.14. $dist(D,G) \le f(a,b,c,d)$,
where $f(a,b,c,d) = \min\{(b+c), \frac{(ad-bc)}{a+b}, \frac{1}{2}(|\sqrt{a}-a-b| + |\sqrt{a}-a-c|, |(1-\sqrt{a})^2-d|)\}$

Proof. Let $G = (x,y,z,t)$ be a product graph.

$$dist(D,G) \le dist(D,G) = \frac{1}{2}(|a-x| + |b-y| + |c-z| + |d-t|)$$

So if we can find some graphs $G's$ such that $dist(D,G)$ equal to the three values above respectively, we are done. From the second property of distribution graph, it suggests the way to choose such $G's$.

1. Choose $y = z = 0$, $x = a+b$, and $t = c+d$, then $dist(D,G) = b+c$.

2. Choose $xt = yz$, $x = a$, $y = b$, $z = \frac{(c+d)a}{a+b}$, and $\frac{(c+d)b}{a+b}$, then

$$dist(D,G) = \frac{1}{2}(|c - \frac{(c+d)a}{a+b}| + |d - \frac{(c+d)b}{a+b}| = \frac{(ad-bc)}{a+b}$$

3. Choose $x = a$, $y = z = \sqrt{a}-a$, and $t = (1-\sqrt{a})^2$, then

$$dist(D,G) = \frac{1}{2}(|\sqrt{a}-a-b| + |\sqrt{a}-a-c|, |(1-\sqrt{a})^2-d|)$$

□

Claim A.15. $m \le \max\limits_{a,b,c,d} f(a,b,c,d)$

Claim A.16. $\max\limits_{a,b,c,d} f(a,b,c,d) = \sqrt{5}-2$

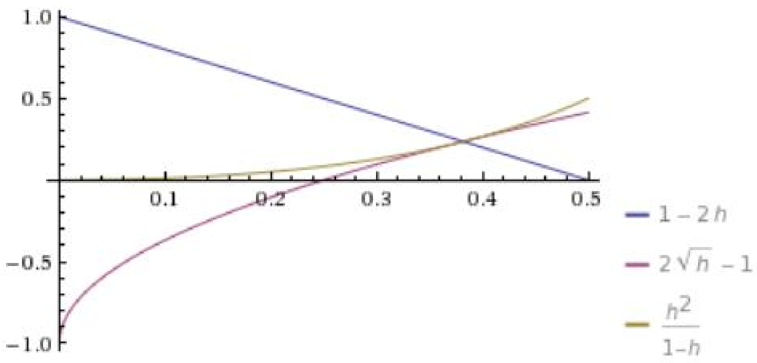
Proof. 1. If $b(1-d) \ge ad$, then $\frac{(ad-bc)}{a+b} = \frac{(ad-bc)}{1-c-d} \le \frac{ad}{1-d} \le \frac{(\frac{a+d}{2})^2}{1-(\frac{a+d}{2})}$. Then

$$\max_{a,b,c,d} f(a,b,c,d) \le \max\min\{(1-2h), \frac{h^2}{1-h}\}$$

where $h = (a+d)/2$

2. If $b(1-d) = ba + b^2 + bc \le ad$, suppose $\max\min\{(1-2h), \frac{h^2}{1-h}\} > \sqrt{5}-2$. Then $\sqrt{a} \ge a+b \ge a+c$ since $a \ge (a+b)^2 \Leftrightarrow a(a+b+c+d) \ge a^2+2ab+b^2 \Leftrightarrow ac+ad \ge ab+b^2$, which is true. So $d \ge (1-\sqrt{a})^2$. Thus,

$$\begin{aligned}
&\frac{1}{2}(|\sqrt{a}-a-b| + |\sqrt{a}-a-c|, |(1-\sqrt{a})^2-d|) \\
&= \frac{1}{2}(2\sqrt{a}-2a-(1-a-d)+d-(1-\sqrt{a})^2 \\
&= d-(1-\sqrt{a})^2 \\
&= (\sqrt{d}+\sqrt{a}-1)(1+\sqrt{d}-\sqrt{a}) \\
&\le \sqrt{d}+\sqrt{a}-1 \\
&\le \sqrt{2(a+d)}-1
\end{aligned}$$



Hence,

$$\max_{a,b,c,d} f(a,b,c,d) \leq \max\min\{1-2h, 2\sqrt{h}-1\}\}$$

From the graph, we can see that

$$\max\min\{1-2h, 2\sqrt{h}-1, \frac{h^2}{1-h}\} = \sqrt{5}-2$$

when $a = d = \frac{3-\sqrt{5}}{2}$, $b = \sqrt{5}-2$, and $c = 0$ □