

PURDUE UNIVERSITY

CS 699

SPRING 2016

---

# Research Thesis

---

*Author:*

Hai NGUYEN

*Instructor:*

Prof. Hemanta MAJI

March 9, 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  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ . 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  $S \subseteq [n] = \{1, 2, \dots, n\}$ , 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,

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

□

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

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

*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  $\Delta$  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  $\chi_S$  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  $\chi_S$  is an orthonormal set. Also, there are  $2^n$  different  $\chi_S$ . Hence, the set of all  $\chi_S$  must be an orthonormal basis for the space of all real-valued functions on  $\{0,1\}^n$ . □

The set of all  $\chi_S$  is called 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:

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

**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:

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

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  $\chi_S$  forms an orthonormal basis,

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

Hence,

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

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

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

This also implies:

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

## 1.6 Convolution

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

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

*Then*

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

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

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

**Claim 1.16.**  $\langle v_1 \rangle^\perp \cap \dots \cap \langle v_k \rangle^\perp = \langle v_1, \dots, v_k \rangle^\perp = V^\perp$

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.**  $f = f_1 \oplus f_2 \oplus \dots \oplus f_k$

Hence,

$$\widehat{f}(S) = N^{k-1} \widehat{f}_1(S) \dots \widehat{f}_k(S)$$

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

If  $S \notin V^\perp$ , then there exists 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  $X = (x_0, x_1, \dots, x_{N-1})$  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  $\sum_{i=0}^{N-1} x_i^2 = N \cdot \|X\|_2^2$ .*

**Definition 2.3** (Flat Distribution). A probability distribution function  $f: \{0, 1\}^n \rightarrow (0, 1)$  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  $\alpha$ -flat distribution can be written as the sum of  $\beta$ -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  $\sum_i \alpha_i = 1$ .*

*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.  $\square$

**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,

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

$\square$

**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  $\alpha$ -bias if  $\widehat{\mathcal{D}}(S) \leq \frac{\alpha}{N}$ .

**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,

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

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  $G_0 \sim 1 \times \frac{n}{2}$ ,  $G \sim \frac{n}{2} \times n$ , and let  $X$  be a uniform distribution over  $\{0, 1\}^{\frac{n}{2}}$ . Then*

$$SD\{(XG_0, XG \oplus M, G_0, G), (U, XG \oplus M, G_0, G)\} \leq \dots$$

*Proof.* For convenience, let  $A = (XG_0, XG \oplus M | G_0, G)$  and  $B = (U, XG \oplus M | G_0, G)$ .

**Claim 2.13.** For any distributions  $C, D$ ,

$$\widehat{(C, D)}(S) = \dots, \text{ where } S = S_C S_D$$

*Proof.*

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

□

**Claim 2.14.** For any  $S \subseteq [n+1]$ ,  $\hat{A}(S) = \hat{B}(S)$  if  $S_1 = 0$ , and  $\hat{B}(S) = 0$  if  $S_1 = 1$ .

Back to the problem,

$$\begin{aligned}
& SD\{(XG_0, XG \oplus M, G_0, G), (U, XG \oplus M, G_0, G)\} \\
&= \mathbb{E}_{G_0, G} [SD(A, B)] \\
&= \mathbb{E}_{G_0, G} \left[ \frac{1}{2} \sum_{i,j} |A(i) - B(i)| \right] \\
&\leq \frac{1}{2} \mathbb{E}_{G_0, G} \left[ \sqrt{2N \cdot \sum_i ((A(i) - B(i))^2)} \right] \\
&\leq \frac{\sqrt{2N}}{2} \sqrt{\mathbb{E}_{G_0, G} \left[ \sum_i (A(i) - B(i))^2 \right]} \\
&= \frac{\sqrt{2N}}{2} \sqrt{\sum_S \mathbb{E}_{G_0, G} \left[ (\hat{A}(S) - \hat{B}(S))^2 \right]} \\
&= \frac{\sqrt{2N}}{2} \sqrt{\sum_{S_1=1} \mathbb{E}_{G_0, G} \left[ \hat{A}(S)^2 \right]}
\end{aligned}$$

□

## Chapter 3

# Bourgain's Extractor



# 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  $V^\perp = \{x \in \{0, 1\}^n \mid x \cdot v = 0 \forall v \in V\}$ .

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

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

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

**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  $V = \{\mathbf{0}, v\}$ . Suppose the number of  $v_i = 1$  is  $t$ , then the number of  $x$  such that  $x \cdot v = 0$  is  $\sum_{i:2|t-i} \binom{n}{i} 2^{n-t} = 2^{t-1} \cdot 2^{n-t} = 2^{n-1}$  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  $V = \langle v_1, v_2, \dots, v_k \rangle$ ,  $S_1 = \langle v_1, v_2, \dots, v_{k-1} \rangle$ , 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  $v_1, v_2, \dots, v_{k-1}$ . Thus,  $t = n - k$ . So  $|V^\perp| = 2^{n-k}$ .  $\square$

## 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*

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

*Proof.*

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

□

## 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.6.** We say  $\psi : G \rightarrow \mathbb{C}$  is a character if  $\psi$  is a homomorphism.

**Definition A.7.** We say a map  $e : G \times G \rightarrow \mathbb{C}$  is a bilinear map if it is a homomorphism in each variable.

**Theorem A.8.** *For every abelian group  $G$ , there exists a symmetric non-degenerate bilinear  $e : G \times G \rightarrow \mathbb{C}$*

### A.3.2 Dual of a finite Abelian Group

**Theorem A.9.** *Every finite abelian group  $G$  is isomorphic to its character group  $G^\wedge$*



## A.4 Product Graph

**Definition A.10.**  $D = (a, b, c, d)$  is a  $2 \times 2$  distribution graph if and only if

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

**Definition A.11.**  $G = (x, y, z, t)$  is a  $2 \times 2$  product graph if and only if

1.  $G$  is a distribution graph,
2.  $xt = yz$ , or  $x = t = 0$ , or  $y = z = 0$ .

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

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

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

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

**Claim A.12.**  $\operatorname{dist}(D, \mathbb{G}) \leq f(a, b, c, d)$ ,  
 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.

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

So if we can find some graphs  $G$ 's such that  $\operatorname{dist}(D, G)$  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.

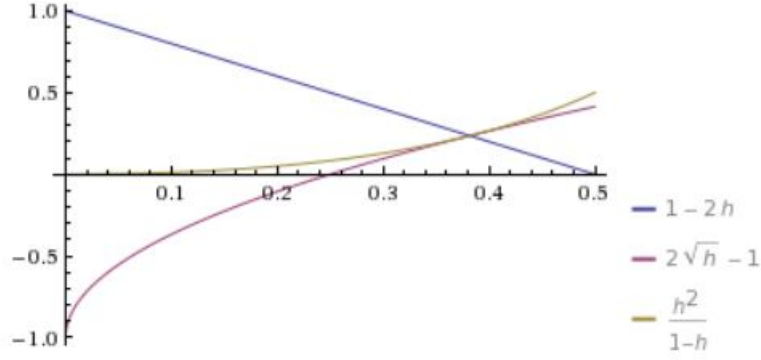
1. Choose  $y = z = 0$ ,  $x = a + b$ , and  $t = c + d$ , then  $\operatorname{dist}(D, G) = b + c$ .
2. Choose  $xt = yz$ ,  $x = a$ ,  $y = b$ ,  $z = \frac{(c+d)a}{a+b}$ , and  $\frac{(c+d)b}{a+b}$ , then

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

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

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

□



**Claim A.13.**  $m \leq \max_{a,b,c,d} f(a, b, c, d)$

**Claim A.14.**  $\max_{a,b,c,d} f(a, b, c, d) = \sqrt{5} - 2$

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

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

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

2. If  $b(1-d) = ba + b^2 + bc \leq ad$ , suppose  $\max \min\{(1-2h), \frac{h^2}{1-h}\} > \sqrt{5} - 2$ .

Then  $\sqrt{a} \geq a+b \geq a+c$  since  $a \geq (a+b)^2 \Leftrightarrow a(a+b+c+d) \geq a^2 + 2ab + b^2 \Leftrightarrow ac + ad \geq ab + b^2$ , which is true. So  $d \geq (1-\sqrt{a})^2$ . Thus,

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

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  $c = 0$

□