Purdue University

CS 699

Spring 2016

Research Thesis

Author:
Hai Nguyen

Instructor:
Prof. Hemanta Maji

February 24, 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 $S, T \subseteq [n]$,*

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

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 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 (Fourier transform function)**.** For each $S \subseteq [n]$, we define the Fourier transform of f as following:

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

Theorem 1.7. *The mapping $\mathcal{F} : f \to \widehat{f}$ is linear.*

Proof. This follows from the properties of inner product.
$\langle af + bg, \chi_S \rangle = \langle af, \chi_S \rangle + \langle bg, \chi_S \rangle = a\langle f, \chi_S \rangle + b\langle g, \chi_s \rangle = a\widehat{f} + b\widehat{g}$ □

The linear map $\mathcal{F} : f \to \widehat{f}$ is called the *Fourier transform* .

Theorem 1.8. *The linear map $\mathcal{F}$ is a bijection.*

Proof. Since the set of χ_S forms an orthonormal basis,

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

Suppose $\mathcal{F}(f_1) = \mathcal{F}(f_2)$, i.e, $\widehat{f_1}(S) = \widehat{f_2}(S)$ for every S, then it is followed from equation (1.1) that $f_1 = f_2$. So $\mathcal{F}$ is injective.
Also, equation (1.1) implies that for every $\widehat{f}$, there exists a function $f = \sum_S \widehat{f}(S)\chi_S$ such that $\mathcal{F}(f) = \widehat{f}$, which means that $\mathcal{F}$ is surjective.
Thus, $\mathcal{F}$ is a bijection as derived □

1.5 Convolution

Definition 1.9. 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 1.10. *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 1.11. *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 1.12. *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 1.13. $V = \langle v_1 \rangle \oplus \langle v_2 \rangle \oplus ... \oplus \langle v_k \rangle$

Claim 1.14. $\langle v_1 \rangle^\perp \cap ... \cap \langle v_k \rangle^\perp = \langle v_1, ..., v_k \rangle^\perp = V^\perp$

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 1.15. $f = f_1 \oplus f_2 \oplus ... \oplus f_k$

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, $\widehat{f}(S) = 0$ □

1.6 Parseval's Identity

Because the χ_S form an orthonormal basis, we have the following equality:

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

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:

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

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.

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

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 $f\colon \{0,1\}^n \to (0,1)$ 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 $\sum_i \alpha_i = 1$.*

Proof. Let S be the set of all the probability distributions X with $H_\infty(X) \geq k$, then S is a compact convex polytope in $\mathbb{R}^N$. The set of all 2^k-flat distributions is the set of vertices (extreme points) of S. Since S is a compact convex set, S equals to the set of all convex combinations of its vertices. Hence, every distribution X with $H_\infty(X) \geq k$ can be written as a convex combination of k-flat distributions.

□

Theorem 2.6. *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 2.7. *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 2.8 (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 2.9. *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 2.10. *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,

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

$\square$

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

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 $|V^\perp| = 2^{n-k}$. $\square$