1 Basics

Our main object of study are boolean functions $f : \{0, 1\}^n \to \{0, 1\}$. Examples are:

- In circuit complexity, these correspond to circuits which take $n$ input bits and output a one bit answer.
- In learning theory, these correspond to concepts with $n$ binary attributes.
- In combinatorics, $f$ can be thought of as the characteristic function of $A \subset \{0, 1\}^n$. That is, $f(x) = 1$ iff $x \in A$.
- In coding theory, $f$ can be the characteristic function of a code $C \subset \{0, 1\}^n$.
- In graph theory, if $n = \binom{m}{2}$ then $f$ describes an undirected graph on $m$ nodes.
- In social choice theory, $f$ corresponds to a voting system with $n$ participants and two possible outcomes, 0 and 1.

Notation. In different scenarios we will identify 0, 1 with True,False or with 1,−1. The most useful representation often will be $f : \{-1, 1\}^n \to \{-1, 1\}$.

Hamming cube. The domain of $f$ is called the hamming cube (other names: boolean hypercube, $n$-cube, boolean cube, discrete cube). The hamming distance between $x, y \in \{0, 1\}^n$ is

$$\Delta(x, y) = \#\{i : x_i \neq y_i\}.$$ 

Fourier analysis is often useful when the hamming distance arises in a problem.
2 Fourier decomposition

The fourier decomposition of \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \) is its expression as a multilinear polynomial (that is, the degree of each variable is 0 or 1 in each monomial).

**Example 2.1.** Let \( f(x_1, x_2) = \max(x_1, x_2) \). That is, \( f(1, 1) = 1, f(1, -1) = 1, f(-1, 1) = 1, f(-1, -1) = -1 \). Its Fourier decomposition is

\[
f(x_1, x_2) = \frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_1x_2.
\]

For \( x \in \{-1, 1\}^n \) and \( S \subseteq [n] \), let \( x^S = \prod_{i \in S} x_i \).

**Lemma 2.2.** Any function \( f : \{-1, 1\}^n \rightarrow \mathbb{R} \) can be expressed as

\[
f(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} f_S \cdot x^S.
\]

**Proof.** Let’s first prove this for point functions, \( 1_a(x) = 1 \) for \( a \in \{-1, 1\}^n \). We can express

\[
1_a(x) = \prod_{i=1}^{n} \frac{a_i x_i + 1}{2}.
\]

In particular, \( \hat{1_a}(S) = a^S = \prod_{i \in S} a_i \). For a general \( f \), we can express

\[
f(x) = \sum_{a \in \{-1, 1\}^n} f(a)1_a(x)
\]

and apply the above decomposition to each function \( 1_a \).

We will later see that the Fourier decomposition of \( f \) is unique. Thus, we define the Fourier coefficients of \( f \) as \( \hat{f}(S) = f_S \) as above.

Another way to view boolean functions are as \( f : \mathbb{F}_2^n \rightarrow \{-1, 1\} \), where \( \mathbb{F}_2 = \{0, 1\} \) is the field of two elements.

**Definition 2.3** (Characters). The characters of \( \mathbb{F}_2^n \) are the functions \( \chi_S : \mathbb{F}_2^n \rightarrow \{-1, 1\} \) for \( S \subseteq [n] \), defined as

\[
\chi_S(x) = \prod_{i \in S} (-1)^{x_i} = (-1)^{\sum_{i \in S} x_i}.
\]

We will sometime also write \( \chi_a(x) \) for \( a \in \{0, 1\}^n \), where we define \( \chi_a(x) = \chi_S(x) \) for \( S = \{i : a_i = 1\} \). With this notation,

\[
\chi_a(x) = (-1)^{\langle a, x \rangle}.
\]

Therefore, the Fourier decomposition of \( f : \mathbb{F}_2^n \rightarrow \mathbb{R} \) is given by

\[
f(x) = \sum_{S \subseteq [n]} \hat{f}(S)\chi_S(x).
\]

Observe that for \( \chi_S : \mathbb{F}_2^n \rightarrow \{-1, 1\} \) we have \( \chi_S(x + y) = \chi_S(x)\chi_S(y) \).
3 Orthogonality of characters

Let's return to the notation of \( f : \{-1,1\}^n \rightarrow \mathbb{R} \). In this notation the characters are \( \chi_S(x) = x^S \). Given two functions \( f, g : \{-1,1\}^n \rightarrow \mathbb{R} \), there is a natural inner product defined on them,

\[
\langle f, g \rangle = 2^{-n} \sum_{x \in \{-1,1\}^n} f(x)g(x).
\]

We will use the following notation from now onwards. For \( f : \{-1,1\}^n \rightarrow \mathbb{R} \) we denote \( \mathbb{E}[f] = 2^{-n} \sum_{x \in \{-1,1\}^n} f(x) \) its average. Thus \( \langle f, g \rangle = \mathbb{E}[fg] \).

Claim 3.1. Let \( S, T \subseteq [n] \). Then \( \chi_{S \Delta T} = \chi_S \chi_T \), where \( S \Delta T \) is the symmetric difference of \( S \) and \( T \).

Proof. \( \chi_S(x)\chi_T(x) = \prod_{i \in S} x_i \prod_{i \in T} x_i = \prod_{i \in S \Delta T} x_i \cdot \prod_{i \in S \cap T} x_i^2 = \prod_{i \in S \Delta T} x_i = \chi_{S \Delta T}(x) \).

Claim 3.2. if \( S \subseteq [n] \) is nonempty then \( \mathbb{E}[\chi_S] = 0 \).

Proof. \( \mathbb{E}[\chi_S] = 2^{-n} \sum_{x \in \{-1,1\}^n} \prod_{i \in S} x_i = 2^{-n} \prod_{i \in S} \left( \sum_{x_i \in \{-1,1\}} x_i \right) = 0 \).

Lemma 3.3. The characters \( \{\chi_S(x) : S \subseteq [n]\} \) form an orthogonal basis for the functions \( \mathbb{F}_2^n \rightarrow \mathbb{R} \).

Proof. Clearly \( \langle \chi_S, \chi_S \rangle = \mathbb{E}[\chi_S^2] = 1 \), since \( \chi_S(x) \in \{-1,1\} \) for all \( x \). If \( S, T \) are distinct then \( \langle \chi_S, \chi_T \rangle = \mathbb{E}[\chi_S \chi_T] = \mathbb{E}[\chi_{S \Delta T}] = 0 \) as \( S \Delta T \) is nonempty. This implies that the characters are linearly independent as functions \( \mathbb{F}_2^n \rightarrow \mathbb{R} \). Indeed, if

\[
p(x) = \sum p_S \chi_S(x) = 0 \quad \forall x \in \{-1,1\}^n
\]

then

\[
p_S = \langle \chi_S, p \rangle = 0.
\]

As we have \( 2^n \) linearly independent functions in a linear space of dimension \( 2^n \), they must form a basis.

Corollary 3.4. The Fourier decomposition is unique.

Proof. Assume \( f : \{-1,1\}^n \rightarrow \mathbb{R} \) can be expressed in two different ways as a multilinear polynomials, say \( p_1(x) \) and \( p_2(x) \). Then \( p(x) = p_1(x) - p_2(x) \) is a nonzero multilinear polynomial, which evaluates to 0 on \( \{-1,1\}^n \). Assume that \( p(x) = \sum_S p_S \chi_S(x) \). As we just saw, \( p_S = \langle \chi_S, p \rangle = 0 \) for all \( S \), which is a contradiction.

Corollary 3.5 (Fourier inversion formula). Let \( f : \{-1,1\}^n \rightarrow \mathbb{R} \). Assume its Fourier decomposition is \( f(x) = \sum_S \hat{f}(S) \chi_S(x) \). Then

\[
\hat{f}(S) = \langle f, \chi_S \rangle = \mathbb{E}[f(x) \prod_{i \in S} x_i].
\]
4 Useful identities

Claim 4.1. \( E[f] = \hat{f}(\emptyset) \).

Proof. \( E[f] = \langle f, 1 \rangle = \langle f, \chi_\emptyset \rangle = \hat{f}(\emptyset) \). \qed

Lemma 4.2 (Parseval’s identity). Let \( f : \{-1,1\}^n \to \mathbb{R} \). Then

\[ E[f^2] = \sum_S \hat{f}(S)^2. \]

In particular, if \( f : \{-1,1\}^n \to \{-1,1\} \) then \( \sum_S \hat{f}(S)^2 = 1 \).

Proof. Let \( f(x) = \sum_S \hat{f}(S)\chi_S(x) \) be the Fourier decomposition of \( f \). Then

\[ f^2(x) = \sum_{S,T} \hat{f}(S)\hat{f}(T)\chi_S(x)\chi_T(x) = \sum_{S,T} \hat{f}(S)\hat{f}(T)\chi_{S\Delta T}(x) \]

and

\[ E[f^2] = E\left[ \sum_{S,T} \hat{f}(S)\hat{f}(T)\chi_{S\Delta T}(x) \right] = \sum_{S,T} \hat{f}(S)\hat{f}(T)E[\chi_{S\Delta T}(x)] = \sum_S \hat{f}(S)^2. \] \qed

Lemma 4.3 (Plancherel’s identity). Let \( f, g : \{-1,1\}^n \to \mathbb{R} \). Then

\[ \langle f, g \rangle = \sum_S \hat{f}(S)\hat{g}(S). \]

Proof. Same as for Parseval’s identity. \qed

The convolution of two functions \( f, g : \mathbb{F}_2^n \to \mathbb{R} \) is the function \( f * g : \mathbb{F}_2^n \to \mathbb{R} \) defined as

\[ (f * g)(x) = \mathbb{E}_{y \in \mathbb{F}_2^n} f(x + y)g(y) \]

Lemma 4.4. \( \hat{f * g}(S) = \hat{f}(S)\hat{g}(S) \).

Proof. Let \( f(x) = \sum_S \hat{f}(S)\chi_S(x) \) and \( g(x) = \sum_S \hat{g}(S)\chi_S(x) \). Then

\[ (f * g)(x) = \mathbb{E}_y \sum_S \hat{f}(S)\chi_S(y) \sum_T \hat{g}(T)\chi_T(x + y) \]

\[ = \sum_{S,T} \hat{f}(S)\hat{g}(T)\mathbb{E}_y[\chi_S(y)\chi_T(x + y)] \]

\[ = \sum_{S,T} \hat{f}(S)\hat{g}(T)\chi_T(x)\mathbb{E}_y[\chi_{S\Delta T}(y)] \]

\[ = \sum_S \hat{f}(S)\hat{g}(S)\chi_S(x). \] \qed
5 BLR linearity test

Let $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ be an unknown function. The goal is to see if $f$ is a linear function, namely if $f(x) = \sum a_i x_i$ for some $a \in \mathbb{F}_2^n$.

The first observation is that being a linear function can be characterized locally.

**Claim 5.1.** $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ is a linear function iff $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{F}_2^n$.

**Proof.** If $f$ is linear then clearly $f(x + y) = f(x) + f(y)$. For the other direction, define $a_i = f(e_i)$, where $e_i$ is the unit vector with 1 in coordinate $i$ and 0 everywhere else. We claim that $f(x) = \sum a_i x_i$. We prove this by induction on the hamming weight of $x$, $|x| = \#\{i : x_i \neq 0\}$. First, note that as $f(0) = f(0 + 0) = f(0) + f(0) = 0$ then $f(0) = 0$. Also by definition it holds if $|x| = 1$. So, assume $|x| > 1$, and let $f$ be such that $x_j = 1$. Let $y = x + e_j$ where $|y| = |x| - 1$. Then

$$f(x) = f(e_j) + f(y) = a_j + \sum a_i y_i = \sum a_i x_i.$$  

This raises the following intriguing question: assume that $f(x + y) = f(x) + f(y)$ for most pairs $x, y$. Does that mean that $f$ is “close” in some sense to a linear function. Otherwise put, can we distinguish between linear functions, and functions “far” from linear functions, simply by picking a few random pairs $x, y$ and testing if $f(x + y) = f(x) + f(y)$. The answer turns out to be yes, and this result started the field of property testing, where a few local queries can identify global properties of functions (such as being linear).

Let $f, g : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$. Their relative distance if the faction of coordinates where they differ,

$$\text{dist}(f, g) = \Pr_{x \in \mathbb{F}_2^n} [f(x) \neq g(x)] = 2^{-n} \sum_{x \in \mathbb{F}_2^n} 1_{f(x) \neq g(x)}.$$  

The distance from linearity of $f$ is its distance from the closest linear function,

$$\text{dist}(f, \text{linear}) = \min_{g \text{ linear}} \text{dist}(f, g).$$  

**Theorem 5.2.** Let $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$. Let $p = \Pr_{x,y \in \mathbb{F}_2^n} [f(x + y) = f(x) + f(y)]$. Then

(i) If $\text{dist}(f, \text{linear}) \leq \varepsilon$ then $p \geq 1 - 3\varepsilon$. In particular, if $f$ is linear then $p = 1$.

(ii) If $\text{dist}(f, \text{linear}) \geq \varepsilon$ then $p \leq 1 - \varepsilon$.

**Proof.** The proof of (i) is straightforward. Let $g$ be a linear function such that $\text{dist}(f, g) \leq \varepsilon$. Let $A = \{x : f(x) = g(x)\}$, where $|A| \geq (1 - \varepsilon)2^n$. Let $x, y \in \mathbb{F}_2^n$ be chosen randomly. Observe that if all $x, y, x + y \in A$ then $f(x + y) = g(x + y) = g(x) + g(y) = f(x) + f(y)$. Thus

$$1 - p \leq \Pr[x \notin A \vee y \notin A \vee x + y \notin A] \leq \Pr[x \notin A] + \Pr[y \notin A] + \Pr[x + y \notin A] \leq 3\varepsilon.$$  

5
We now move to prove (ii). Define $F : \mathbb{F}_2^n \to \{-1, 1\}$ by $F(x) = (-1)^{f(x)}$. We study $F(x)F(y)F(x + y)$. Note that this expression evaluates to 1 exactly when $f(x + y) = f(x) + f(y)$, and otherwise evaluates to $-1$. Thus

$$E_{x,y}[F(x)F(y)F(x + y)] = p - (1 - p) = 2p - 1.$$ 

Next, we decompose $F$ to its Fourier coefficients. Let $F(x) = \sum_S \hat{F}(S)\chi_S(x)$. Then

\[
\begin{align*}
F(x)F(y)F(x + y) &= \sum_{S,T,R} \hat{F}(S)\hat{F}(T)\hat{F}(R)\chi_S(x)\chi_T(y)\chi_R(x + y) \\
&= \sum_{S,T,R} \hat{F}(S)\hat{F}(T)\hat{F}(R)\chi_S(x)\chi_T(y)\chi_R(x)\chi_R(y) \\
&= \sum_{S,T,R} \hat{F}(S)\hat{F}(T)\hat{F}(R)\chi_{S\Delta R}(x)\chi_{T\Delta R}(y)
\end{align*}
\]

and

\[
E_{x,y}[F(x)F(y)F(x + y)] = \sum_{S,T,R} \hat{F}(S)\hat{F}(T)\hat{F}(R)E_x[\chi_{S\Delta R}(x)]E_y[\chi_{T\Delta R}(y)] = \sum_S \hat{F}(S)^3.
\]

Next, we bound each $\hat{F}(S)$. Let $g(x) = \sum_{i\in S} x_i$ be a linear function. We have

$$\hat{F}(S) = \mathbb{E}[F(x)\chi_S(x)] = \Pr[f = g] - \Pr[f \neq g] \leq 1 - 2\varepsilon.$$ 

Thus we can bound

$$E_{x,y}[F(x)F(y)F(x + y)] = \sum_S \hat{F}(S)^3 \leq (1 - 2\varepsilon)\sum_S \hat{F}(S)^2 = 1 - 2\varepsilon,$$

where we applied Parseval’s identity to conclude that $\sum_S \hat{F}(S)^2 = 1$. Combining everything we get that

$$2p - 1 = E_{x,y}[F(x)F(y)F(x + y)] \leq 1 - 2\varepsilon$$

and hence $p \leq 1 - \varepsilon$. \qed

Consider the following randomized test (the BLR test).

1. Pick uniformly and randomly $x, y \in \mathbb{F}_2^n$.
2. Query the value of $f(x), f(y), f(x + y)$.
3. If $f(x + y) = f(x) + f(y)$ then accept $f$, otherwise reject $f$. 

6
Corollary 5.3. Let $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$. Then

(i) If $f$ is linear then the basic BLR test always accepts $f$.

(ii) If $\text{dist}(f, \text{linear}) \geq \varepsilon$ then the basic BLR test rejects $f$ with probability at least $\varepsilon$.

In order to amplify the rejection probability, we will simply repeat the basic BLR test independently $1/\varepsilon$ times.

Lemma 5.4. Fix $\varepsilon > 0$. Let $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2$. Consider an algorithm which runs the basic BLR test $m = 1/\varepsilon$ times independently, and accepts $f$ only if all the tests accepted $f$. Then

(i) If $f$ is linear then the test always accepts $f$.

(ii) If $\text{dist}(f, \text{linear}) \geq \varepsilon$ then the test rejects $f$ with probability at least $1/e$.

Proof. (i) is clear. Assume then that $\text{dist}(f, \text{linear}) \geq \varepsilon$. By the Theorem, the basic BLR test rejects $f$ with probability at least $\varepsilon$. So, the probability that all $m$ tests accept $f$ is at most $(1 - \varepsilon)^m \leq 1/e$. 

$\square$