

## Lecture 1

*Lecturer: Irit Dinur**Scribe: Shay Mozes*

# 1 Introduction

The main object of study in this course are boolean functions  $f : \{0,1\}^n \rightarrow \{0,1\}$ , which are very natural objects in a discrete, digital world. The usual course of mathematical (undergrad) education first introduces continuous functions, and discrete ones are introduced much later. Sometimes one needs to make a decision based on some value, a number  $x \in [0,1]$ . For example, in a voting situation where  $x$  percent of the population favors certain action. Since in the end one needs to decide yes or no, one has to perform a "rounding", moving from a continuous function to a discrete one. The further  $x$  is from 0 or 1, the worse the "rounding error". This tension between continuous and discrete that is found in real life also manifests itself when analyzing Boolean functions.

## 1.1 Aspects of Boolean Functions that we will study

- Range is  $\{0,1\}$  as opposed to  $\mathbb{R}$ , or  $[0,1]$ . These functions correspond to "decisions", not "valuations"
- Domain is  $\{0,1\}^n$  is the product of  $n$  *independent* bits. The fact that the domain is a product of  $n$  independent spaces, be it bits or some other space, will be very important.
- Probability space (implicit) over the domain. It is convenient to view as the product of  $n$  individual probability spaces instead of just the uniform distribution over  $n$ -bit strings.

We will not be talking in this course about worst case behavior, but mostly consider average case (the probability is over the inputs). When measuring norms/inner-products one always gets the expectation value.

- Graph Structure over  $\{0,1\}^n$  (metric). The Hamming Cube  $H_n = (\{0,1\}^n, E)$ , where for  $x, y \in \{0,1\}^n$ ,  $(x, y) \in E$  iff  $x$  and  $y$  differ in exactly one bit. The Hamming distance corresponds to shortest paths metric in  $H_n$ . We will be interested in, e.g., how a function behaves in relations to the metric (noise-stable functions are those that are not sensitive to small changes under the metric). Another example is the dependency of a function on a single specific variable, which corresponds to moving along an edge of the graph. Other connections include the fact that a boolean function is a 2-coloring of  $H_n$ . We will see relations to natural questions on graphs. E.g., cuts, expansion, etc...

The course will cover a set of tools. We will look at cool problems/applications using those tools, but the goal is mostly introducing and studying the tools rather than the specific applications.

# 2 The Fourier Basis

A boolean function  $f : \{-1,1\}^n \rightarrow \{-1,1\}$  can be thought of as a vector (a truth table)

$$f \in \mathbb{R}^{2^n} = \underbrace{\mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \cdots \mathbb{R}^2}_{n \text{ times}}.$$

This is perhaps not the most natural way to think about a Boolean function, but this perspective will turn out very useful.

**Definition 1** (*tensor product*) For vector spaces  $U = \mathbb{R}^{d_1}, V = \mathbb{R}^{d_2}$  and vectors  $u \in U, v \in V$ , the tensor product of  $u$  and  $v$  is the vector  $u \otimes v \in \mathbb{R}^{d_1 \times d_2}$  given by

$$(u \otimes v)_{ij} = u_i v_j.$$

The tensor product of  $U$  and  $V$  is

$$U \otimes V = \text{span}\{u \otimes v : u \in U, v \in V\}.$$

Example:  $\mathbb{1} = (1, 1), \chi = (1, -1)$

$$\mathbb{1} \otimes \chi = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$$

$$\chi \otimes \chi = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$\chi \otimes \mathbb{1} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$$

$$\mathbb{1} \otimes \mathbb{1} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and generally  $\underbrace{\mathbb{1} \otimes \chi \mathbb{1} \otimes \dots \otimes \mathbb{1}}_{n \text{ times}} \in \mathbb{R}^{2^n} = \mathbb{R}^2 \otimes \mathbb{R}^2 \otimes \dots \otimes \mathbb{R}^2$ . Note that we don't have to write a tensor product in matrix form (very inconvenient in high dimensions...)

**Claim 2**  $\mathbb{R}^2 \otimes \dots \otimes \mathbb{R}^2 = \mathbb{R}^{2^n}$

By definition,  $\text{LHS} \subseteq \mathbb{R}^{2^n}$ . We need to show  $\supseteq$ . We will show that LHS contains a set of  $2^n$  independent vectors. This will be the Fourier basis.

We will use the following notation. It is useful to think of the vectors  $\mathbb{1}$  and  $\chi$  as functions. We will use  $x$  to denote the coordinate.  $x = 1$  is the first coordinate and  $x = -1$  is the second coordinate. We therefore have:

$$\mathbb{1}(x) \equiv 1 \text{ and } \chi(x) \equiv x.$$

$$\mathbb{1} \otimes \chi(x_1, x_2) = \mathbb{1}(x_1)\chi(x_2) = x_2$$

$$\chi \otimes \mathbb{1}(x_1, x_2) = \chi(x_1)\mathbb{1}(x_2) = x_1$$

$$\chi \otimes \chi(x_1, x_2) = \chi(x_1)\chi(x_2) = x_1 x_2$$

**Definition 3** (*Fourier Basis*) For  $S \subseteq [n]$ , let  $\chi_S \stackrel{\text{def}}{=} \bigotimes_1^n S_i$ , where  $S_i = \begin{cases} \mathbb{1} & i \notin S \\ \chi & i \in S \end{cases}$

In other words, for every  $S, x_1, \dots, x_n$ ,

$$\chi_S(x_1, \dots, x_n) = \prod_{i \in S} x_i.$$

The functions  $\{\chi_S\}_S$  are called the Fourier Basis Functions or the Fourier Characters.

**Definition 4** (*Inner Product on  $\mathbb{R}^N$* ) For  $f, g : \{-1, 1\}^n \rightarrow \mathbb{R}$ , define the standard inner product by

$$\langle f, g \rangle \stackrel{\text{def}}{=} \mathbb{E}_{x \in \{-1, 1\}^n} [f(x)g(x)] = \frac{1}{2^n} \sum_{x \in \{-1, 1\}^n} f(x)g(x).$$

**Claim 5** If  $S \neq T$  then  $\chi_S \perp \chi_T$  (i.e.,  $\langle \chi_S, \chi_T \rangle = 0$ ).

**Proof** For  $n = 1$  then clearly  $\langle \chi, \mathbb{1} \rangle = \frac{1}{2}(1 \cdot 1 + 1 \cdot (-1)) = 0$ .

In general, if  $f_1 \otimes f_2, g_1 \otimes g_2 \in \mathbb{R}^{d_1 \times d_2}$ , then, using  $x$  as indices for  $f_1, g_1$  and  $y$  for  $f_2, g_2$ ,

$$\begin{aligned} \langle f_1 \otimes f_2, g_1 \otimes g_2 \rangle &= \mathbb{E}_{x,y} f_1 \otimes f_2(x,y) \cdot g_1 \otimes g_2(x,y) \\ &= \mathbb{E}_{x,y} f_1(x) f_2(y) g_1(x) g_2(y) \\ &= \mathbb{E}_x f_1(x) g_1(x) \mathbb{E}_y f_2(y) g_2(y) \\ &= \langle f_1, g_1 \rangle \langle f_2, g_2 \rangle \end{aligned}$$

Therefore, for  $S \neq T$ ,  $\langle \chi_S, \chi_T \rangle = \prod_{i=1}^n \langle \chi_{S_i}, \chi_{T_i} \rangle = 0$ , since at least one term in the product involves  $\langle \chi, \mathbb{1} \rangle$  or  $\langle \mathbb{1}, \chi \rangle$  which are 0. ■

**Definition 6** (*Fourier Coefficients*) For  $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ , the Fourier Coefficient of  $f$  at  $S$  is

$$\hat{f}_S \stackrel{\text{def}}{=} \langle f, \chi_S \rangle = \mathbb{E}_{x \in \{-1, 1\}^n} [f(x) \chi_S(x)].$$

**Claim 7**  $f = \sum \hat{f}_S \cdot \chi_S$

**Proof**  $f = \alpha_S \chi_S$  because  $\chi_S$  form a basis of  $\mathbb{R}^{2^n}$ . But

$$\hat{f}_S = \langle f, \chi_S \rangle = \langle \sum \alpha_{S'} \chi_{S'}, \chi_S \rangle = \alpha_S.$$

■

**Claim 8 (Parseval)** Given  $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ ,  $\sum \hat{f}_S^2 = 1$ .

**Proof** Since the range of  $f$  is  $\{-1, 1\}$ ,  $f(x)^2 = 1$  for all  $x$ , so  $\langle f, f \rangle = \mathbb{E}_x [f(x)f(x)] = 1$ . But, using the Fourier representation of  $f$ ,  $\langle f, f \rangle = \langle \sum_S \hat{f}_S \chi_S, \sum_S \hat{f}_S \chi_S \rangle = \sum_S \hat{f}_S^2$ . ■

The choice of basis  $\mathbb{1}, \chi$  seems a little arbitrary at this point. We will later see some reasons for this. Another natural basis to have started with is the standard basis  $(0, 1), (1, 0)$  for  $\mathbb{R}^2$ . When tensorized, this yields the so-called *standard* (or computational) basis for  $\mathbb{R}^{2^n}$ . Many of the results that we shall see in this course will relate the natural description of a function in the standard basis to the description through the Fourier basis. For example, the “address” function, takes the first  $\log n$  bits, treats them as an index and outputs the bit at that index. While it is easy to describe in the standard basis, the Fourier representation looks much less intuitive.

### 3 Linearity Testing

Linearity testing started the field of property testing and has been a very important influence and starting points for several other research directions.

Boolean function can be written as functions over  $GF_2$ ,  $f : GF_2^n \rightarrow GF_2$ .

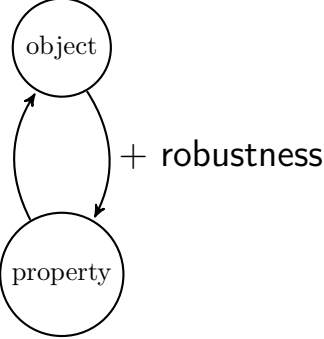
**Definition 9** (*Linear functions*)  $f$  is a linear function if there exist  $\alpha_1, \dots, \alpha_n \in GF_2$  such that for all  $\vec{a} = (a_1 \dots a_n) \in \{0, 1\}^n$ ,

$$f(a_1 \dots a_n) = \sum_{i=1}^n \alpha_i a_i \pmod{2}.$$

A linear function is just a parity of a subset of its bits.

**Fact 10**  $f$  is linear if and only if  $\forall a, b, f(a) + f(b) = f(a + b)$

We are so used to this fact that we often treat the latter as the definition for linear functions. Let's explore what happens to this equivalence in the presence of noise or error. One direction is clear: if a linear function is corrupted on 1% of the entries, does the above fact still (mostly) hold? Yes, by union bound,  $f(a) + f(b) = f(a + b)$  still holds with probability at least 0.97.



The other direction requires more work. Suppose we know that  $f(a) + f(b) = f(a + b)$  holds for at least 99% of the choices of  $a, b$ . Is  $f$  necessarily close to a linear function? In other words, is there a small (say, 3 percent) part of the input space on which  $f$  can be changed to make it into a linear function?

Asking these kinds of questions helps understand which properties of an object are intrinsic and robustly so. It is an example of a current trend of trying to understand if well-known equalities still hold in the presence of noise, and how much noise can they tolerate.

**Theorem 11** *If  $\text{Prob}_{a,b} [f(a) + f(b) = f(a + b)] \geq 1 - \epsilon$  then there exists a linear function  $g$  such that  $\text{Prob}_a [f(a) = g(a)] \geq 1 - \epsilon$*

To prove this theorem we need to find a linear function that is close to  $f$ . It is instructive to first try to think whether there can be a function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  very far from any linear function yet  $\text{Prob}_{a,b} [f(a) + f(b) = f(a + b)] \geq 0.97$ .

Consider any  $c \in \{0, 1\}^n$ . For each choice of  $a, b$  such that  $b = c + a$ , we get a “guess” for  $f(c)$ . We can try to show that by taking the majority over these guesses, the resulting function is linear and close to  $f$  with high probability. This is essentially the proof given by Blum, Luby and Rubinfeld.

The proof we will give is different. We first move from  $GF_2$  to  $\{-1, 1\}$  by  $0 \leftarrow 1$  and  $1 \leftarrow -1$ . I.e.,  $x_i = (-1)^{a_i}$ . If  $p : \{0, 1\}^n \rightarrow \{0, 1\}$  is linear (i.e.,  $p = \sum \alpha_i a_i \pmod{2}$ ), define  $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$  by

$$f(x_1, \dots, x_n) = f((-1)^{a_1}, \dots, (-1)^{a_n}) = (-1)^{p(a, \dots, a_n)}.$$

Since  $p$  is linear,  $f(x_1, \dots, x_n) = (-1)^{\sum \alpha_i a_i} = \prod (-1)^{\alpha_i a_i} = \chi_S$ , where  $S = \{i : \alpha_i = 1\}$ . That is,  $f$  is a tensor function! (This is why the Fourier basis elements are sometimes called the parity functions).

Let us rewrite the theorem in multiplicative notation. For vectors  $x, y$  of the same dimensions we will use the notation  $x \odot y$  to denote the coordinate-wise product of  $x$  and  $y$ . I.e.,  $(x \odot y)_i = x_i y_i$ .

**Theorem 12 (BLR)** *Let  $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$  such that  $\text{Prob}_{x,y} [f(x)f(y) = f(x \odot y)] \geq 1 - \epsilon$ . Then there exists  $S$  such that  $\text{Prob}_x [f(x) = \chi_S(x)] \geq 1 - \epsilon$ .*

**Proof** Since the range is  $\{-1, 1\}$ ,  $f(x)f(y) = f(x \odot y)$  implies  $f(x)f(y)f(x \odot y) = 1$ . We want to find some  $S$  such that  $f$  is close to  $\chi_S$ . It is natural to look in the Fourier representation of  $f$ . First, let us write the hypothesis for  $f$  in analytical terms,

$$\begin{aligned} \mathbb{E}_{x,y \in \{-1, 1\}^n} [f(x)f(y)f(x \odot y)] &= \mathbb{E}_{x,y} \left[ \left( \sum_S \hat{f}_S \chi_S(x) \right) \left( \sum_T \hat{f}_T \chi_T(y) \right) \left( \sum_W \hat{f}_W \chi_W(x \odot y) \right) \right] \\ &= \mathbb{E}_{x,y} \left[ \sum_{S,T,W} \hat{f}_S \chi_S(x) \hat{f}_T \chi_T(y) \hat{f}_W \chi_W(x \odot y) \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{S,T,W} \hat{f}_S \hat{f}_T \hat{f}_W \mathbb{E}_x [\chi_S(x) \chi_W(x)] \mathbb{E}_y [\chi_T(y) \chi_W(y)] \\
&= \sum_{S=T=W} \hat{f}_S \hat{f}_T \hat{f}_W \\
&= \sum_S \hat{f}_S^3 \\
&\leq \max_s \hat{f}_S \sum_S \hat{f}_S^2 \\
&= \max_s \hat{f}_S
\end{aligned} \tag{1}$$

Here we used the fact that  $\chi_W(x \odot y) = \prod_{i \in W} (x \odot y)_i = \prod_{i \in W} x_i \prod_{i \in W} y_i = \chi_W(x) \chi_W(y)$ , and Parseval's theorem. Now, by definition,

$$\hat{f}_S = \mathbb{E}_x f(x) \chi_S(x) = \text{Prob}_x [f(x) = \chi_S(x)] - \text{Prob}_x [f(x) \neq \chi_S(x)] = 2\text{Prob}_x [f(x) = \chi_S(x)] - 1.$$

Similarly,

$$\mathbb{E}_{x,y} f(x) f(y) f(x \odot y) = 2\text{Prob}_x [f(x) f(y) = f(x \odot y)] - 1.$$

By Eq. 1 there exists  $S_0$  such that  $\hat{f}_{S_0} \geq \mathbb{E}[f(x) f(y) f(x \odot y)]$ . Hence,

$$\text{Prob}_x [f(x) = \chi_{S_0}(x)] \geq \text{Prob}_{x,y} [f(x) f(y) = f(x \odot y)] \geq 1 - \epsilon.$$

■

Our proof is related to that of BLR in factorizing  $\mathbb{E}_{x,y} f(x) f(y) f(x \odot y)$  into  $\mathbb{E}_x f(x) \underbrace{\mathbb{E}_y f(y) f(x \odot y)}_{g(x)}$ . BLR

“round”  $g(x)$  by taking the majority, while our proof uses the expectation value. We actually get a stronger result than BLR. In the large error regime (say  $\epsilon = 0.49$ ) we still get that  $f$  is “close” to some linear function (albeit with probability 0.51). This is significantly different than just a random function.