Notes on Fourier Analysis of Boolean Functions

Anil Ada

1 Notation

Random variables are denoted with boldface letters, not necessarily capital. If $\mathbf{x}$ is a random variable and μ a distribution, $\mathbf{x} \sim \mu$ means that $\mathbf{x}$ is distributed according to μ. The notation $\mathbf{E}[\cdot]$ and $\mathbf{Pr}[\cdot]$ is used for expectation and probability respectively. When the random variable(s) and the distribution(s) are clear from the context, the expectations and the probabilities do not have any subscripts, e.g. $\mathbf{E}[f(\mathbf{x})]$. If the distribution is clear but we would like to explicitly point out the random variables, we put the random variables as subscript, e.g. $\mathbf{E}_{\mathbf{x}}[f(\mathbf{x})]$. We also sometimes choose to make the distribution explicit in this notation, e.g. $\mathbf{E}_{\mathbf{x}\sim\mu}[f(\mathbf{x})]$. The uniform distribution is always denoted by U and the underlying set will always be clear from the context.

We use the notation $[n]$ to denote the set $\{1, 2, \ldots, n\}$. When x is a bit string, $|x|$ denotes the number of 1's in x, i.e., the Hamming weight of x. If $x \in \{0,1\}^n$, x_i denotes the ith bit of x.

A function is called a *boolean function* if it has the form $f : \mathcal{S} \to \{0,1\}$, where $\mathcal{S}$ is some set. For convenience, we will often define the range of a boolean function as $\{1,-1\}$ rather than $\{0,1\}$ with the understanding that -1 corresponds to 1 and 1 corresponds to 0. In other words, $f(x)$ is represented as $(-1)^{f(x)}$. A function $f : \{0,1\}^n \to \{1,-1\}$ is called *symmetric*if the output depends only on $|x|$, the Hamming weight of the input. In other words, the output does not change if we permute the input bits.

2 Fourier Analysis of Boolean Functions Basics

The study of boolean functions $f : \{0,1\}^n \to \{0,1\}$ is central to complexity theory and combinatorics as objects of interest in these areas can often be represented as boolean functions. Fourier analysis of boolean functions provides some of the strongest tools in this study with applications to graph theory, circuit complexity, communication complexity, hardness of approximation, machine learning, etc.

As before, for convenience, we will view the range as $\{1,-1\}$ rather than $\{0,1\}$. The main idea behind Fourier analysis of boolean functions is very simple. We are interested in studying the set of boolean functions $\mathcal{B} = \{f : \{0,1\}^n \to \{1,-1\}\}$. This set by itself does not have much structure and therefore is not easy to reason about. On the other hand, vector spaces have a lot of structure and we understand them very well. Therefore a natural thing to do is to view $\mathcal{B}$ as residing in a vector space, and the natural candidate is the vector space of real valued functions $\mathcal{V} = \{p : \{0,1\}^n \to \mathbb{R}\}$. This is a 2^n-dimensional vector space over the reals. Furthermore, we can turn $\mathcal{V}$ into an inner product space by defining an appropriate inner product: for $p, q \in \mathcal{V}$, define

$$\langle p, q \rangle \stackrel{\text{def}}{=} \mathbf{E}\left[p(\mathbf{x})q(\mathbf{x})\right],$$

where the expectation is with respect to the uniform distribution over $\{0,1\}^n$. Thus we can equivalently write

$$\langle p, q \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} p(x)q(x).$$

This absolute value of the inner product is often called the *correlation* between p and q because when p and q are boolean functions, the inner product really measures how well p and q are correlated. For boolean functions f and g, we define the **correlation** as

$$\mathrm{Cor}(f, g) \stackrel{\text{def}}{=} \left|\mathbf{Pr}\left[f(\mathbf{x}) = g(\mathbf{x})\right] - \mathbf{Pr}\left[f(\mathbf{x}) \neq g(\mathbf{x})\right]\right|.$$

Observe that this quantity is always between 0 and 1. It is 1 when $f(x) = g(x)$ for all x or $f(x) = -g(x)$ for all x. It is 0 when $f(x)$ and $g(x)$ agree on exactly half the points x (i.e., knowing $f(x)$ for a random x tells us nothing about $g(x)$). Since f and g are ± 1-valued functions, a moment's observation shows that the correlation can be alternatively written as

$$\mathrm{Cor}(f, g) = \left|\mathbf{E}\left[f(\mathbf{x})g(\mathbf{x})\right]\right| = |\langle f, g \rangle|.$$

More generally, for a probability distribution μ over $\{0,1\}^n$, we define the correlation of f and g under μ as

$$\mathrm{Cor}_\mu(f, g) \stackrel{\text{def}}{=} \left|\mathbf{E}_{\mathbf{x}\sim\mu}\left[f(\mathbf{x})g(\mathbf{x})\right]\right| = \left|\sum_x f(x)g(x)\mu(x)\right|.$$

Now that we have an inner product, we can hope to define a *useful* orthonormal basis. The Fourier basis consists of the following functions. For each $S \subseteq [n]$, define $\chi_S : \{0,1\}^n \to \{1, -1\}$ as

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

In other words, the value of $\chi_S(x)$ is the parity of the variables in S, where -1 means the parity is odd and 1 means the parity is even. These functions are often called *characters* and in our setting we have 2^n of them. It is straightforward to verify that

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

So we conclude that the set of characters form on orthonormal basis for the vector space $\mathcal{V}$. This means that we can write every $p \in \mathcal{V}$ as a linear combination of the characters:

$$p(x) = \sum_{S \subseteq [n]} \widehat{p}(S) \chi_S(x).$$

Here, $\widehat{p}(S) \in \mathbb{R}$ denotes the coefficient corresponding to χ_S, and these coefficients are called the **Fourier coefficients**. This way of expanding p as a linear combination of the characters is called the **Fourier expansion** of p. Since the characters form on orthonormal basis, it follows that $\widehat{p}(S) = \langle p, \chi_S \rangle$. We will call the set of Fourier coefficients of p the **Fourier spectrum** of p.

Remark. It is worth noting that Fourier analysis can be applied more generally in the setting of $\mathcal{V} = \{f : G \to \mathbb{C}\}$, where G denotes an Abelian group (we view $\{0,1\}^n$ as $\mathbb{F}_2^n$ so we are in the special case of $G = \mathbb{F}_2^n$). A character $\chi : G \to \mathbb{C}$ is any function that satisfies $\chi(gh) = \chi(g)\chi(h)$ for all $g, h \in G$ (when $G = \mathbb{F}_2^n$, the parity functions χ_S are the only functions with this property). The set of all characters form an orthonormal basis for $\mathcal{V}$ with respect to the inner product $\langle p, q \rangle = \mathbf{E}\left[p(\mathbf{x})\overline{q(\mathbf{x})}\right]$. Therefore all functions in $\mathcal{V}$ can be written as a linear combination of the characters.

The elements of $\mathcal{V}$ are often referred to as *polynomials* (which is why we chose to use the notation p). The reason for this is as follows. If we view the domain of p as $\{1, -1\}^n$ rather than $\{0,1\}^n$, then observe that the characters take the form

$$\chi_S(x) = \prod_{i \in S} x_i.$$

That is, each character is a *multilinear*[1] monomial and the Fourier expansion of p is simply a multilinear polynomial representation of the function. There is no real difference between the two representations and we will stick with the domain $\{0,1\}^n$.

Since $\mathcal{B} \subset \mathcal{V}$, every boolean function $f : \{0,1\}^n \to \{1,-1\}$ also has a Fourier expansion:

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

In essence, Fourier analysis of boolean functions is the study of boolean functions by looking at the information of how well the function correlates with different parity functions. This point of view turns out to be quite fruitful and we will see several applications of it in this thesis.

Let us now dive a little deeper and try to explore interesting features of Fourier analysis. We start with the most fundamental and essential fact, often called Parseval's Identity, which forms the bridge between the usual representation of a function in terms of the values $\{p(x) \mid x \in \{0,1\}^n\}$ and the Fourier representation in terms of the Fourier

[1]Multilinear means that each variable has exponent 0 or 1.

coefficients $\{\widehat{p}(S) \mid S \subseteq [n]\}$. It states that the inner product we defined for $\mathcal{V}$ (i.e., the expected value of the product of the functions) is the usual dot product of the Fourier coefficients.

Fact 2.1 (Parseval's Identity)**.** *For $p, q \in \mathcal{V}$,*

$$\langle p, q \rangle = \sum_{S \subseteq [n]} \widehat{p}(S)\widehat{q}(S).$$

Remark. Sometimes the above fact is called Plancherel's Theorem and the special case of $p = q$ is called Parseval's Identity. For convenience we call the general case Parseval's Identity.

Proof. The proof simply substitutes the Fourier expansion of p and q in the definition of the inner product and then uses the orthonormality of the characters:

$$\begin{aligned}\langle p, q \rangle = \mathbf{E}\left[p(\mathbf{x})q(\mathbf{x})\right] &= \mathbf{E}\left[\sum_{S \subseteq [n]} \widehat{p}(S)\chi_S(\mathbf{x}) \sum_{T \subseteq [n]} \widehat{q}(T)\chi_T(\mathbf{x})\right] \\ &= \sum_{S,T} \widehat{p}(S)\widehat{q}(T)\mathbf{E}\left[\chi_S(\mathbf{x})\chi_T(\mathbf{x})\right] = \sum_{S} \widehat{p}(S)\widehat{q}(S).\end{aligned}$$

□

A basic corollary of this fact is that for boolean functions, $\sum_S \widehat{f}(S)^2 = 1$. This is easy to see by substituting $p = q = f$ in Parseval's Identity. This allows us to view the squares of the Fourier coefficients of a boolean function as a probability distribution over the sets $S \subseteq [n]$. We will call this the **Fourier distribution**. In many different settings, how close this distribution is to the uniform distribution determines how *complex* a function is. There are of course various ways to measure how close a distribution is to the uniform distribution and which one to use depends on the context and application.

The previous paragraph in fact outlines a general theme about how Fourier analysis is used in computational complexity theory. In many different settings, the *hardness* of a function exposes itself in the function's Fourier expansion. In other words, different analytic measures associated with the Fourier coefficients of f can be good approximations to how *complex* the function is e.g., in communication complexity, circuit complexity, learning theory etc. Let us now define some of these useful measures.

The **degree** of a function p is the degree of the multilinear polynomial representation of p. In other words, degree of p, denoted $\deg(p)$, is defined as $\max\{|S| : \widehat{p}(S) \neq 0\}$. The **monomial complexity** of p is the number of monomials in its polynomial representation, i.e., $|\{S \mid \widehat{p}(S) \neq 0\}|$. We denote it by $\mathrm{mon}(p)$.

The usual L_p norms are defined as (to avoid confusion we are using f instead of p to denote a general element of $\mathcal{V}$):

$$\|f\|_p = \mathbf{E}\left[|f(\mathbf{x})|^p\right]^{1/p}.$$

With respect to the Fourier coefficients, we define

$$\|\widehat{f}\|_p = \left(\sum_S |\widehat{f}(S)|^p \right)^{1/p}.$$

Recall that Parseval's Identity implies $\|f\|_2 = \|\widehat{f}\|_2$ and for boolean functions this quantity is 1. We characterize this situation by saying that the total L_2 mass of a boolean function is 1. Other interesting L_p norms are the Fourier L_1 norm and the Fourier L_∞ norm. For a boolean function we have

$$1 \le \|\widehat{f}\|_1 \le 2^{n/2}.$$

The lower bound follows from the fact that $\|\widehat{f}\|_1 \ge \|\widehat{f}\|_2$ and the upper bound follows from the Cauchy-Schwarz inequality and $\|\widehat{f}\|_2 = 1$. Also we have

$$\frac{1}{2^{n/2}} \le \|\widehat{f}\|_\infty \le 1.$$

The lower bound follows from the fact that $\|\widehat{f}\|_2^2 \le 2^n \max_S \widehat{f}(S)^2$ and the upper bound follows from $\|\widehat{f}\|_\infty \le \|\widehat{f}\|_2$. The Fourier L_1 and L_∞ norms are measures of how close the Fourier distribution is to the uniform distribution. In fact the Fourier L_1 norm corresponds to the Rényi entropy of order 1/2 of the Fourier distribution:[2]

$$H_{1/2}[\widehat{f}^2] = 2 \log \left(\sum_S |\widehat{f}(S)| \right) = 2 \log \|\widehat{f}\|_1.$$

The Fourier L_∞ norm corresponds to the min-entropy:

$$H_\infty[\widehat{f}^2] = -\log \|\widehat{f}\|_\infty^2.$$

At the one extreme, we have the constant function $f \equiv 1$, with $\|\widehat{f}\|_1 = \|\widehat{f}\|_\infty = 1$. On the other extreme, the *inner-product* function satisfies $|\widehat{f}(S)| = 1/2^{n/2}$ for all S, i.e., its Fourier distribution is uniform. Therefore $\|\widehat{f}\|_\infty = 1/2^{n/2}$ and $\|\widehat{f}\|_1 = 2^{n/2}$.

Let us move on to other useful measures that can serve as the complexity of a boolean function. We say that $p \in \mathcal{V}$ **sign represents** a boolean function f if $p(x)f(x) > 0$ for all x, in other words, $f(x) = \text{sign}(p(x))$ for all x. The **sign degree** of f, denoted $\deg_\pm(f)$, is the minimum degree of a function p that sign represents f. Similarly, the **sign monomial complexity** of f, denoted $\text{mon}_\pm(f)$, is the minimum monomial complexity of a function p that sign represents f. As an example, first consider the *majority* function. Observe that $\text{MAJ}(x) = \text{sign}((-1)^{x_1} + \cdots + (-1)^{x_n} - 0.5)$ and so $\deg_\pm(\text{MAJ}) = 1$ and $\text{mon}_\pm(\text{MAJ}) \le n+1$. On the other hand, it is quite straightforward

[2] For $\alpha > 0$ and $\alpha \neq 1$, the Rényi entropy of order α is defined as $H_\alpha(\mathbf{X}) = \frac{1}{1-\alpha} \log \left(\sum_{x \in \mathcal{X}} \mathbf{Pr}[\mathbf{X} = x]^\alpha \right)$.

to show that the *parity* function satisfies $\deg_{\pm}(\text{PAR}) = n$. Let p sign represent PAR. Then $\langle p, \text{PAR}\rangle > 0$ by definition of sign representation. Because $\text{PAR} = \chi_{[n]}$ and the characters are orthogonal, we have $\langle p, \text{PAR}\rangle = 0$ for any p with $\deg(p) \leq n-1$. Therefore the function p that sign represents PAR must have degree n. It is also not too difficult to show that $\text{mon}_{\pm}(\text{IP}) \geq 2^{n/2}$. In fact a classic result of Bruck [Bru90] shows that $\text{mon}_{\pm}(f) \geq \|\widehat{f}\|_{\infty}^{-1}$.

A function p **ϵ-approximates** f if for all x, $|p(x) - f(x)| \leq \epsilon$. In other words, p approximates f within ϵ in the infinity norm: $\|f - p\|_{\infty} \leq \epsilon$. We can define **$\epsilon$-approximate degree** (**ϵ-approximate monomial complexity**, **ϵ-approximate p-norm**) as the minimum degree (monomial complexity, p-norm) of a function that ϵ-approximates f. We denote these quantities by $\deg_{\epsilon}(f), \text{mon}_{\epsilon}(f)$ and $\|\widehat{f}\|_{p,\epsilon}$. We think of ϵ as a fixed constant in the range $[0, 1]$ such as $1/3$. A classic result of Paturi [Pat92], that has found many applications in theoretical computer science, characterizes the approximate degree of all symmetric functions. Let $t_0(f) \in [\lfloor n/2 \rfloor]$ and $t_1 \in [\lceil n/2 \rceil]$ be the minimum integers such that $f(i) = f(i+1)$ for all $i \in [t_0(f), n - t_1(f)]$.

Theorem 2.2 ([Pat92])**.** *For $f : \{0,1\}^n \to \{1, -1\}$ a symmetric function,*

$$\deg_{1/3}(f) = \Theta\left(\sqrt{n(t_0(f) + t_1(f))}\right).$$

All these measures $\deg(f)$, $\text{mon}(f)$, $\|\widehat{f}\|_p$, $\deg_{\pm}(f)$, $\text{mon}_{\pm}(f)$, $\deg_{\epsilon}(f)$, $\text{mon}_{\epsilon}(f)$, $\|\widehat{f}\|_{p,\epsilon}$ can serve as a reasonable measure of complexity of f depending on the particular context. Note that by definition we have

$$\deg_{\pm}(f) \leq \deg_{\epsilon}(f) \leq \deg(f),$$

$$\text{mon}_{\pm}(f) \leq \text{mon}_{\epsilon}(f) \leq \text{mon}(f),$$

and

$$\|\widehat{f}\|_{p,\epsilon} \leq \|\widehat{f}\|_p.$$

3 Noise Stability

In this section we will introduce a very important concept in Fourier analysis of boolean functions: the noise operator and noise stability. In many different situations, the noise operator serves as the crucial connection between the combinatorial properties of a boolean function and its Fourier properties.

We begin by defining the noise operator. For $x \in \{0,1\}^n$ and $\rho \in [0, 1]$, we say that $\mathbf{y}$ is a ρ-noisy copy of x, denoted $\mathbf{y} \sim_{\rho} x$, if $\mathbf{y}$ is such that for each $i \in [n]$ independently, we have:

$$\mathbf{y}_i = \begin{cases} x_i & \text{with probability } \rho, \\ 0 & \text{with probability } \frac{1-\rho}{2}, \\ 1 & \text{with probability } \frac{1-\rho}{2}. \end{cases}$$

We also write $\mathbf{y}_i \sim_\rho x_i$ when the coordinates have the above relation. We write $\mathbf{y} \sim_\rho \mathbf{x}$ when $\mathbf{x}$ is uniformly distributed over $\{0,1\}^n$ and $\mathbf{y}$ is then chosen to be a ρ-noisy copy of $\mathbf{x}$. Note that we have symmetry: $\mathbf{y}$ has uniform distribution over $\{0,1\}^n$ and $\mathbf{x} \sim_\rho \mathbf{y}$.

For $f : \{0,1\}^n \to \mathbb{R}$, we define the **noise operator** T_ρ to be such that

$$T_\rho f(x) = \mathbf{E}_{\mathbf{y} \sim_\rho x}\left[f(\mathbf{y})\right].$$

It is easy to check that T_ρ is linear in the sense that

$$T_\rho(f + cg) = T_\rho f + cT_\rho g.$$

Now let us see how T_ρ affects a function's Fourier expansion.

Proposition 3.1. *Let* $f : \{0,1\}^n \to \mathbb{R}$*. Then,*

$$T_\rho f = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S) \chi_S.$$

Proof. Since T_ρ is a linear operator, it suffices to show that $T_\rho \chi_S = \rho^{|S|}\chi_S$, which is quite straightforward:

$$\begin{aligned} T_\rho \chi_S(x) &= \mathbf{E}_{\mathbf{y} \sim_\rho x}\left[\chi_S(\mathbf{y})\right] = \mathbf{E}_{\mathbf{y} \sim_\rho x}\left[\prod_{i \in S}(-1)^{\mathbf{y}_i}\right] = \prod_{i \in S} \mathbf{E}_{\mathbf{y}_i \sim_\rho x_i}\left[(-1)^{\mathbf{y}_i}\right] \\ &= \prod_{i \in S} \rho(-1)^{x_i} = \rho^{|S|}\chi_S. \end{aligned}$$

□

This proposition shows that the noise operator dampens the high degree Fourier coefficients and the dampening increases exponentially with the degree.

With regards to boolean functions, our main interest will be in how sensitive a function is when noise is applied to its input. To measure this, we look at the correlation of the function with its noisy version. More formally, define the **noise stability** of a function $f : \{0,1\}^n \to \mathbb{R}$ to be

$$\mathrm{Stab}_\rho(f) \stackrel{\text{def}}{=} \mathbf{E}_{\mathbf{y} \sim_\rho \mathbf{x}}\left[f(\mathbf{x})f(\mathbf{y})\right].$$

For boolean functions, this corresponds to

$$\mathbf{Pr}\left[f(\mathbf{x}) = f(\mathbf{y})\right] - \mathbf{Pr}\left[f(\mathbf{x}) \neq f(\mathbf{y})\right].$$

By the definition of the noise operator, we can equivalently write

$$\mathrm{Stab}_\rho(f) = \mathbf{E}_{\mathbf{x} \sim U}\left[f(\mathbf{x})T_\rho f(\mathbf{x})\right] = \langle f, T_\rho f \rangle.$$

Using Parseval's identity and Proposition 3.1, we see that the noise stability of a function has a clean Fourier formula:

$$\mathrm{Stab}_\rho(f) = \langle f, T_\rho f \rangle = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S)^2.$$

This in particular shows that the noise stability of a function is always non-negative, which is not immediately obvious from the original definition of noise stability. Also observe that the above formula implies that noise stable functions must have significant Fourier weight on the low degree coefficients. This intuitively makes sense too since high degree characters are very noise sensitive.

References

[Bru90] Jehoshua Bruck. Harmonic analysis of polynomial threshold functions. *SIAM Journal of Discrete Mathematics*, 3:168–177, 1990.

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric Boolean functions (preliminary version). In *Proceedings of ACM Symposium on Theory of Computing*, pages 468–474, 1992.