A Short Proof of a Fourier Theorem

Christino Tamon*
Department of Computer Science
University of Calgary
2500 University Dr NW
Calgary AB CANADA T2N 1N4

Abstract

A theorem of Kahn, Kalai, and Linial [2] stated that the average sensitivity of a Boolean function is equal to the weighted sum of its Fourier power spectrum. The purpose of this note is to provide a short proof of this result that is based on a cross correlation Fourier identity. Furthermore we generalize this to product distributions and derive an alternative proof of a theorem in [1].

Keywords: Computational Complexity; Fourier Transform; Sensitivity.

1 Introduction

In complexity theory, the seminal paper of Kahn, Kalai, and Linial [2] initiated the use of harmonic or Fourier analysis on the hypercube in the study of Boolean functions. Their main interest was on the influences of variables on Boolean functions in relation to coin flipping protocols and weak random sources. Their paper contains a famous Fourier theorem that states the average sensitivity of a Boolean function equals to the weighted sum of its Fourier power spectrum. This theorem has been used in other works, e.g. in relation to the sensitivity of AC^0 functions [3], to the "degree" of Boolean functions [4], to the sensitivity of monotone functions and to the learnability of Boolean functions [1].

In this note we provide a short proof the Fourier theorem that is based on a cross correlation Fourier lemma. To the best of our knowledge, this proof as well as the cross correlation lemma have not appeared explicitly in literature before.

2 Preliminaries

In this section we define some necessary notation and we introduce the basic theory of the discrete Fourier transform for Boolean functions.

*Research supported in part by a Graduate Research Scholarship from the University of Calgary

The set $\{1, 2, \ldots, n\}$ is denoted by $[n]$. The set of all Boolean n-bit vectors is $\{0,1\}^n$. The $\oplus$ symbol stands for the bitwise exclusive-or operation on $\{0,1\}^n$. The vector $e_i \in \{0,1\}^n$ is the unit vector whose coordinates are all zero except at the i-th coordinate where it is 1. The Hamming weight of $a \in \{0,1\}^n$, denoted $|a|$, is the number of 1's in a. We use the APL Iversonian notation [statement] to mean 1 if the statement holds and 0 otherwise.

In most of our discussions a size parameter n will be fixed throughout. Consider the set $\mathcal{F}$ of all real-valued functions mapping $\{0,1\}^n$ to $\Re$, that is also a real vector space of dimension 2^n over $GF(2)$. We will let Boolean functions be real-valued functions whose range is $\{-1, +1\}$, i.e., $f : \{0,1\}^n \to \{-1, +1\}$. Given a probability distribution D over $\{0,1\}^n$ we may define an inner product $\langle ., . \rangle$ based on D as follows. For any two real-valued functions f, g, let

$$\langle f, g \rangle =_{\text{def}} \sum_{x \in \{0,1\}^n} D(x) f(x) g(x) = E_D[fg].$$

A product distribution D over $\{0,1\}^n$ with parameters $\mu_1, \mu_2, \ldots, \mu_n \in (0,1)$ is defined as follows. For each element $x \in \{0,1\}^n$

$$D(x) = \prod_{i:x_i=0} (1 - \mu_i) \prod_{i:x_i=1} \mu_i.$$

Alternatively, one may think of D as setting each coordinate $i \in [n]$ to 1 with probability μ_i and to 0 with probability $1 - \mu_i$, independently of the other coordinates. The uniform distribution is a special case of product distributions where $\mu_i = 0.5$ for all $i \in [n]$. Next we "redefine" the notion of Hamming weight for product distributions (taken from [1]). The *weight* of $a \in \{0,1\}^n$ under D, denoted $||a||_D$, is defined as

$$||a||_D = \prod_{i \in a} \log \frac{1}{\sigma_i}.$$

Given a product distribution $D(\mu_1, \ldots, \mu_n)$ and $a \in \{0,1\}^n$, we define the real-valued function $\phi_a(x)$ over $\{0,1\}^n$ as follows.

$$\phi_a(x) = \prod_{i:i \in a} \frac{\mu_i - x_i}{\sigma_i},$$

where $\sigma_i = \sqrt{\mu_i(1 - \mu_i)}$. Note that $\frac{\mu_i - x_i}{\sigma_i}$ is a standard normal random variable.

Fact 1 *Given a product distribution $D(\mu_1, \ldots, \mu_n)$, the set of functions $\{\phi_a : a \in \{0,1\}^n\}$ is an orthonormal basis of $\mathcal{F}$.*

In fact, the functions ϕ_a also satisfy the property that $\phi_{a \cdot b}(x \cdot y) = \phi_a(x)\phi_b(y)$, where $|a| = |x|$ and $|b| = |y|$. For the uniform distribution U, it is easy to see that $\phi_a(x) = (-1)^{a.x}$, i.e., ϕ_a is the parity function on the subset $a \subseteq [n]$.

In the ensuing discussions we will assume a fixed product distribution $D(\mu_1, \ldots, \mu_n)$. The Fourier coefficient of f at a is defined as $\tilde{f}(a) = \langle f, \phi_a \rangle$.

Fact 2 *For each $f \in \mathcal{F}$ we have $f(x) = \sum_{a \in \{0,1\}^n} \tilde{f}(a)\phi_a(x)$.*

For the uniform distribution $D = U$, we will use the more standard notation $\hat{f}(a)$ for $\tilde{f}(a)$ and $\chi_a(x) = (-1)^{a \cdot x}$ for $\phi_a(x)$.

We now need to define some complexity-theoretic definitions. The sensitivity of a Boolean function f at $a \in \{0,1\}^n$ is defined as $s_a(f) = \sum_{i=1}^n [f(a) \neq f(a \oplus e_i)]$. Of course one may replace the Iversonian $[f(a) \neq f(a \oplus e_i)]$ with $\frac{1}{2}|f(a) - f(a \oplus e_i)|$. The **average sensitivity** of f with respect to a (product) distribution D is defined as

$$s_D(f) =_{\text{def}} E_{a \in D}[s_a(f)].$$

When D is very clear from context, we will write $s(f)$ instead of $s_D(f)$. The influence of variable x_i on a Boolean function f with respect to a product distribution D is defined as

$$I_{D,i}(f) =_{\text{def}} \Pr_{x \in D}[f(x) \neq f(x \oplus e_i)].$$

Fact 3 *For any Boolean function f, $s(f) = \sum_{i=1}^n I_{D,i}(f)$.*

We will drop the subscripts D when we are dealing with the uniform distribution U, e.g. $I_i(f) = I_{U,i}(f)$ and $E[X]$ instead of $E_U[X]$. Often we use the restriction notation f_b to denote $f_{x \leftarrow b}$ to mean function f with the variable x restricted to $b \in \{0,1\}$ (in this case the variable x will always be clear from context).

We state now the **cross correlation** Fourier lemma that will be useful in deriving other facts. For ease of exposition we first state the lemma for the uniform distribution case, and in the last section we will generalize it for any product distributions.

Lemma 1 *For any real-valued functions f, g over $\{0,1\}^n$ and for any $y \in \{0,1\}^n$*

$$E_x[f(x)g(x \oplus y)] = \sum_a \hat{f}(a)\hat{g}(a)\chi_a(y).$$

Proof We only need to notice that $\chi_a(x \oplus y) = \chi_a(x)\chi_a(y)$. Then starting from the left-hand side, we use $f(x) = \sum_a \hat{f}(a)\chi_a(x)$ and $g(x \oplus y) = \sum_a \hat{g}(a)\chi_a(x \oplus y)$, apply linearity of expectations, and finally use the orthonormality of the χ_a's. □

A simple but important corollary of the lemma is Parseval's identity.

Corollary 1 *For any real-valued function f over $\{0,1\}^n$ $E[f^2(x)] = \sum_a \hat{f}(a)^2$. Moreover, if f is Boolean then $\sum_a \hat{f}(a)^2 = 1$.*

The expression $\sum_a \hat{f}(a)^2$ is called the power spectrum of f (under the uniform distribution). This shows that the power spectrum of any Boolean function equals 1.

3 A Proof for the Uniform Case

We devote this section and the next to prove the following theorem.

Theorem 1 *[2, 1] For any Boolean function f and any product distribution D*

$$s_D(f) = \sum_{i=1}^n I_{D,i}(f) = \sum_{a \in \{0,1\}^n} ||a||_D \tilde{f}(a)^2.$$

We first present the short proof for the uniform distribution case.

Proof **of Theorem 1 for the uniform case:**
Take $g = f$ and $y = e_i$ in the cross correlation lemma. This gives us

$$E[f(x)f(x \oplus e_i)] = \sum_a \hat{f}(a)^2 \chi_a(e_i) = \sum_{a:a_i=0} \hat{f}(a)^2 - \sum_{a:a_i=1} \hat{f}(a)^2 = 1 - 2 \sum_{a:a_i=1} \hat{f}(a)^2.$$

The last step is by Parseval's identity, $\sum_a \hat{f}(a)^2 = 1$. Now note that $I_i(f) = \frac{1}{2}(1 - E[f(x)f(x \oplus e_i)]$, which implies

$$I_i(f) = \sum_{a:a_i=1} \hat{f}(a)^2.$$

Finally we use $s(f) = \sum_{i=1}^n I_i(f)$ to finish the claim. □

4 Product Distributions

In this section we generalize the main lines from the previous section to product distributions. First we restate the cross correlation lemma in this more general setting.

Lemma 2 *For any Boolean functions f, g, for any $y \in \{0,1\}^n$, and for any product distribution $D = (\mu_i)_{i=1}^n$ over $\{0,1\}^n$ we have*

$$E_D[f(x)g(x \oplus y)] = \sum_{a,b} \tilde{f}(a)\tilde{g}(b)E_D[\phi_a(x)\phi_b(x \oplus y)].$$

We need the following intermediate lemma.

Corollary 2 *[1] For any Boolean functions f and for any product distribution D over $\{0,1\}^n$ we have*

$$E_D[f(x)f(x \oplus e_i)] = 1 - \frac{1}{2\sigma_i^2} \sum_{a:a_i=1} \tilde{f}(a)^2.$$

Proof Let $\Delta = E_D[f(x)f(x \oplus e_i)]$, which by Lemma 2, equals

$$\sum_{a,b} \tilde{f}(a)\tilde{f}(b)E_D[\phi_a(x)\phi_b(x \oplus e_i)].$$

Setting $a = \overline{a} \cdot a_i$ and $b = \overline{b} \cdot b_i$, where $\overline{a}, \overline{b} \in \{0,1\}^{n-1}$ and $a_i, b_i \in \{0,1\}$, we see that

$$E[\phi_a(x)\phi_b(x \oplus e_i)] = (1-\mu_i)E[\phi_{\overline{a}}\phi_{\overline{b}}]\left(\frac{\mu_i}{\sigma_i}\right)^{a_i}\left(\frac{\mu_i - 1}{\sigma_i}\right)^{b_i} + \mu_i E[\phi_{\overline{a}}\phi_{\overline{b}}]\left(\frac{\mu_i - 1}{\sigma_i}\right)^{a_i}\left(\frac{\mu_i}{\sigma_i}\right)^{b_i}.$$

Consider the expression $E_D[\phi_a(x)\phi_b(x \oplus e_i)]$. This expression equals $E[\phi_{\overline{a}}\phi_{\overline{b}}]$ if a_i and b_i are both zero, equals $(-E[\phi_{\overline{a}}\phi_{\overline{b}}])$ if both a_i and b_i are one, equals 0 if $a_i = 1$, $b_i = 0$, and equals $(\frac{\mu_i^2 - (1-\mu_i)^2}{\sigma_i} E[\phi_{\overline{a}}\phi_{\overline{b}}])$ if $a_i = 0$ and $b_i = 1$. Hence by orthonormality of the basis functions ϕ,

$$\begin{aligned}
\Delta &= \sum_{\overline{a},\overline{b}} \left[\tilde{f}(0\overline{a})\tilde{f}(0\overline{b}) - \tilde{f}(1\overline{a})\tilde{f}(1\overline{b}) + \left\{\frac{\mu_i^2 - (1-\mu_i)^2}{\sigma_i}\right\}\tilde{f}(0\overline{a})\tilde{f}(1\overline{b})\right] E[\phi_{\overline{a}}\phi_{\overline{b}}] \\
&= \sum_{\overline{a}} (\tilde{f}(0\overline{a})^2 - \tilde{f}(1\overline{a})^2) + \sum_{\overline{a}} \tilde{f}(0\overline{a})\tilde{f}(1\overline{a}) \left\{\frac{\mu_i^2 - (1-\mu_i)^2}{\sigma_i}\right\}.
\end{aligned}$$

Observing that $\tilde{f}(0\overline{a}) = (1-\mu_i)\tilde{f}_0(\overline{a})+\mu_i\tilde{f}_1(\overline{a})$, and that $\tilde{f}(1\overline{a}) = \sigma_i(\tilde{f}_0(\overline{a})-\tilde{f}_1(\overline{a}))$, we obtain

$$\begin{aligned}
&\sum_{\overline{a}} \tilde{f}(0\overline{a})\tilde{f}(1\overline{a}) \left\{ \frac{\mu_i^2 - (1-\mu_i)^2}{\sigma_i} \right\} \\
&= \sum_{\overline{a}} [(1-\mu_i)\tilde{f}_0(\overline{a}) + \mu_i\tilde{f}_1(\overline{a})][\sigma_i(\tilde{f}_0(\overline{a}) - \tilde{f}_1(\overline{a})] \left\{ \frac{\mu_i^2 - (1-\mu_i)^2}{\sigma_i} \right\} \\
&= [\mu_i^2 - (1-\mu_i)^2] \sum_{\overline{a}} [(1-\mu_i)\tilde{f}_0(\overline{a})^2 - \mu_i\tilde{f}_1(\overline{a})^2 + (2\mu_i - 1)\tilde{f}_0(\overline{a})\tilde{f}_1(\overline{a})] \\
&= (2\mu_i - 1)[(1-2\mu_i) + (2\mu_i - 1)\sum_{\overline{a}} \tilde{f}_0(\overline{a})\tilde{f}_1(\overline{a})].
\end{aligned}$$

But note that $\sum_{\overline{a}} \tilde{f}_0(\overline{a})\tilde{f}_1(\overline{a}) = E[f_0(y)f_1(y)] = E_D[f(x)f(x \oplus e_i)]$. Thus we conclude that

$$\Delta = 1 - 2\sum_{\overline{a}} \tilde{f}(1\overline{a})^2 - (2\mu_i - 1)^2 + (2\mu_i - 1)^2\Delta,$$

which implies the claim. □

Proof **of Theorem 1 for the general case:**
We note that $I_{D,i}(f) = \frac{1}{2}(1 - E_D[f(x)f(x \oplus e_i)])$ and recall that $s_D(f) = \sum_{i=1}^{n} I_{D,i}(f)$. Combining this with Corollary 2 completes the claim. □

References

[1] N. H. Bshouty and C. Tamon. On the Fourier Spectrum of Monotone Functions, in: *Proc. 27th Ann. ACM Symp. on Theory of Computing* (1995) 219-228.

[2] J. Kahn, G. Kalai, and N. Linial. The Influence of Variables on Boolean Functions, in: *Proc. 29th Ann. IEEE Symp. on Foundations of Computer Science* (1988) 68-80.

[3] N. Linial, Y. Mansour, and N. Nisan. Fourier Transform, Constant Depth Circuits, and Learnability, in: *J. of ACM*, **40**:3 (1993) 607-620.

[4] N. Nisan and M. Szegedy. On the Degree of Boolean Functions as Real Polynomials, in: *Proc. 24th Ann. ACM Symp. on Theory of Computing* (1992) 462-467.