Harmonic Analysis of Boolean Functions, and applications in CS

Lecture 3

March 3, 2008

Lecturer: Guy Kindler | Scribe by: Noam Arkind | Updated: April 14, 2008

1 Fourier Basis

1.1 Reminder

Our goal in this class is to find a basis for the space of real-valued functions $\mathbb{R}^{\{-1,1\}^n}$ and explore its properties.

Definition 1 *A Walsh product/character corresponding to a set $S \subseteq [n]$ is defined as*

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

Claim 2 *The set of all characters $\{\chi_S\}_{S \subseteq [n]}$ is a basis for $\mathbb{R}^{\{-1,1\}^n}$.*

Proof We showed last class that $\{\chi_S\}_{S \subseteq [n]}$ spans the canonical basis and hence a basis. Alternatively: Every function can be interpreted by a polynomial $f = \sum_{i=1}^{n} a_i m_i(x)$ (see exercise below), in addition w.l.o.g all m_i are multi-linear ($m_i = \prod_{i \in S} x_i$). ■

Exercise. 1 (The interpolation theorem) *Let $C \subseteq \mathbb{R}^n$ be a finite set and let $f : C \to \mathbb{R}$. Then $\exists g : \mathbb{R}^n \to \mathbb{R}, g = \sum_{i=1}^{n} a_i m_i(x)$, m_i- monomials, s.t. $g(x) = f(x) \forall x \in C$.*

Conclusion. $\forall f : \{\pm 1\}^n \to \mathbb{R}$, $\exists! f = \sum_{s \subseteq [n]} \hat{f}(S)\chi_S$.

It follows that the transformation from the space of functions $f : \{\pm 1\}^n \to \mathbb{R}$ to the set of the coefficients $\hat{f} : P([n]) \to \mathbb{R}$, called the Fourier transform is in fact 1-1 and onto.

2 Some properties of $\{\chi_S\}$

2.1 Shift operator

Definition 3 *The y-shift operator for an element $y \in \{\pm 1\}^n$ is defined over $\mathbb{R}^{\{-1,1\}^n}$ as*

$$\sigma_y f(x) = f(xy).$$

Notation. We denote by σ_i the shift operator σ_{e_i} where $e_i = (1, \ldots, -1, \ldots, 1)$ and -1 in the i-th coordinate.

We now wish to examine how does the shift operator affect characters, and understand better the structure of this basis.

$$\sigma_i \chi_S(x) = \chi_S(xe_i) = \prod_{i \in S} x_i e_i = \begin{cases} \chi_S(x) & i \notin S \\ -\chi_S(X) & i \in S \end{cases} .$$

We can see now that the character is affected by a scalar. This is also true for general shifts:

$$\sigma_y(\chi_S) = \chi_S(xy) = \prod_{i \in S} x_i y_i = (\prod_{i \in S} y_i)(\prod_{i \in S} x_i) = \chi_S(y)\chi_S(x).$$

In other words we can say that χ_S is an eigne vector for the shift operator (for any set S and vector y).

Remark In addition to the previous conclusion, we also got from the computation, that the character χ_S is multiplicative, i.e. $\chi_s(xy) = \chi_s(y)\chi_s(x)$. Another easy result is that the set of characters is closed under multiplication which is clear by the following computation

$$\chi_S(x) \cdot \chi_T(x) = (\prod_{i \in S} x_i)(\prod_{i \in T} x_i) = (\prod_{i \in S \setminus T} x_i)(\prod_{i \in S \cap T} x_i)(\prod_{i \in T \setminus S} x_i)(\prod_{i \in T \cap S} x_i) = \prod_{i \in S \triangle T} x_i = \chi_{S \triangle T}(x).$$

3 Norms, Inner products

3.1 Definitions

Definition 4 *We define the inner product of two functions $f, g : \{\pm 1\} \to \mathbb{R}$ as*

$$\langle f, g \rangle = \mathbb{E}_x[f(x)g(x)].$$

Definition 5 *The induced norm is defined as*

$$\|f\|_2 = \sqrt{\langle f, f \rangle} = \sqrt{\mathbb{E}_x[f(x)^2]}$$

and the corresponding metric is defined by

$$d(f, g) = \|f - g\|.$$

We also call $\|f\|_2^2 = \mathbb{E}_x[f^2(x)]$ the weight of f.

Theorem 6 (Cauchy-Schwartz inequality)

$$|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$$

Theorem 7 (Triangle inequality)

$$\|f+g\|_2 \le \|f\|_2 + \|g\|_2$$

or altenatively $\|\frac{f+g}{2}\|_2 \le \frac{\|f\|_2+\|g\|_2}{2}$ *(namely* $\|\cdot\|$ *is convex).*

Definition 8 (The l_p norm)

$$\|f\|_p = \begin{cases} (\mathbb{E}_x[|f(x)|^p])^{\frac{1}{p}} & 1 \le p < \infty \\ \max_x |f(x)| & p = \infty \end{cases}$$

Remark To show that the l_p norm is indeed a norm and satisfies the triangle inequality is not trivial and requires the Minkowski inequality, but we wont get into it here .

4 Fourier meets inner product

We now explore the behavior of $\{\chi_S\}$ with respect to the inner product we just defined.

4.1 Orthonormality

We observe now that $\{\chi_S\}$ is an orthonormal basis. Recall that we defined $\chi_\emptyset(x) = 1$ and that the characters are unbiased, i.e.

$$\mathbb{E}_x[\chi_s(x)] = \begin{cases} 0 & s \ne \emptyset \\ 1 & s = \emptyset \end{cases} .$$

So now we can compute the inner product of two characters

$$\langle \chi_S, \chi_T \rangle = \mathbb{E}_x[\chi_S(x)\chi_T(x)] = \mathbb{E}_x[\chi_{S\triangle T}(x)] = \begin{cases} 0 & S \ne T \\ 1 & S = T \end{cases} .$$

This computation shows that $\{\chi_S\}$ is indeed an orthonormal basis.
Remark The fact that $\{\chi_S\}$ is an orthonormal basis, is also an alternative proof for the fact that it is indeed a basis.

4.2 Implications From Orthonormality

So far we have:

- $\{\chi_S\}$ is an orthonormal basis that contains multi-linear monomials which are common eigne vectors of all shift operators.
- $\{\chi_S\}$ is closed under multiplication, i.e. $\chi_S\chi_T = \chi_{S\triangle T}$.
- $\{\chi_S\}$ contains multiplicative functions, i.e. $\chi_S(xy) = \chi_S(x)\chi_S(y)$.

From this results we have some immediate corollaries. Let $f, g : \{\pm 1\}^n \to \mathbb{R}$:

Corollary 9 *From the fact that $\{\chi_s\}$ is an orthonormal basis, we know from linear algebra that we can easily compute the coefficients with respect to that basis, the formula for the Fourier coefficient is $\hat{f}(S) = \langle f, \chi_S \rangle$.*

Corollary 10 (Plancharel)

$$\langle f, g \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S) \tag{1}$$

This important formula follows from the following computation

$$\langle f, g \rangle = \langle \sum_{S \subseteq [n]} \hat{f}(S)\chi_S, \sum_{S \subseteq [n]} \hat{g}(T)\chi_T \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(T)\langle \chi_S, \chi_T \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S).$$

Corollary 11 (Parseval)

$$\|f\|^2 = \langle f, f \rangle = \mathbb{E}_x[f^2] = \sum_{S \subseteq [n]} \hat{f}(S) \tag{2}$$

This follows directly from (**??**). *Moreover if f is a boolean function then $\sum_{S \subseteq [n]} \hat{f}^2(S) = 1$.*

Corollary 12 *With this formulas we can express the expectation and the variance in terms of the Fourier coefficients:*

$$\mathbb{E}_x[f(x)] = \mathbb{E}_x[f(x) \cdot 1] =< f, \chi_\emptyset >= \hat{f}(\emptyset)$$
$$\mathbb{V}_x[f(x)] = \mathbb{E}_x[f^2(x)] - (\mathbb{E}[f(x)])^2 = \sum_S \hat{f}^2(S) - \hat{f}^2(\emptyset) = \sum_{S \neq \emptyset} \hat{f}^2(S).$$

4.3 Influence In Terms Of Fourier Coefficients

We can use the results we got so far and express the total influence of a real-valued function in terms of the fourier coefficients. Recall we defined the influence as

$$I_i(f) = \mathbb{E}_{x \setminus i}\big[\mathbb{V}_{x_i}[f(x)]\big]. \tag{3}$$

Remark We use the notation $f(x\backslash i, -1)$ to express f of x where the i-th coordinate is -1.

$$\begin{aligned}
\mathbb{V}_{x_i}[f(x)] &= \mathbb{E}_{x_i}[f^2(x)] - (\mathbb{E}_{x_i}[f(x)])^2 \\
&= \frac{1}{2}f^2(x\backslash i, 1) + \frac{1}{2}f^2(x\backslash i, -1) - (\frac{1}{2}f(x\backslash i, 1) + \frac{1}{2}f(x\backslash i, -1))^2 \\
&= \frac{1}{4}f^2(x\backslash i, 1) + \frac{1}{4}f^2(x\backslash i, -1) - \frac{2}{4}f(x\backslash i, 1)f(x\backslash i, -1) \\
&= \left(\frac{f(x\backslash i, 1) - f(x\backslash i, -1)}{2}\right)^2 \\
&= \mathbb{E}_{x_i}\left[\left(\frac{f(x) - \sigma_i f(x)}{2}\right)^2\right]
\end{aligned}$$

We assign this value in (3) and get

$$\begin{aligned} I_i(f) &= \mathbb{E}_{x\setminus i}\left[\mathbb{E}_{x_i}\left[\left(\frac{f(x)-\sigma_i f(x)}{2}\right)^2\right]\right] = \mathbb{E}_x\left[\left(\frac{f(x)-\sigma_i f(x)}{2}\right)^2\right] \\ &= \left\|\frac{f(x)-\sigma_i f(x)}{2}\right\|_2^2. \end{aligned}$$

Now we can use the fact that the shift operator is linear and compute farther

$$\begin{aligned} \frac{f(x)-\sigma_i f(x)}{2} &= \frac{1}{2}\sum_S \hat{f}(S)\chi_S(x) - \frac{1}{2}\sum_S \hat{f}(S)\sigma_i\chi_S(x) \\ &= \frac{1}{2}\sum_S \hat{f}(S)\chi_S(x) - \frac{1}{2}\left(\sum_{S:i\notin S}\hat{f}(S)\chi_S(x) - \sum_{S:i\in S}\hat{f}(S)\chi_S(x)\right) \\ &= \sum_{S:i\in S}\hat{f}(S)\chi_S(x) \end{aligned}$$

So we have by (2)

$$\begin{aligned} I_i(f) &= \mathbb{E}_{x\setminus i}\big[\mathbb{V}_{x_i}[f(x)]\big] = \left\|\frac{f(x)-\sigma_i f(x)}{2}\right\|_2^2 \\ &= \left\|\sum_{S:i\in S}\hat{f}(S)\chi_S(x)\right\|_2^2 = \sum_{S:i\in S}\hat{f}^2(S). \end{aligned}$$

With this simple formula we can easily express the total influence as

$$I(f) = \sum_i I_i(f) = \sum_{S\subseteq[n]} |S|\hat{f}^2(S) \tag{4}$$

4.4 Applications

We can now use the tools we developed to get a much simpler proof for the theorem from last lecture

$$I(f) = \sum_{S\subseteq[n]} |S|\hat{f}^2(S) \geq \sum_{S\neq\emptyset}\hat{f}^2(S) = \mathbb{V}_x[f(x)] \tag{5}$$

moreover we can farther conclude that if f is boolean and balanced (i.e. $\hat{f}(\emptyset) = \mathbb{E}_x[f(x)] = 0$) then by (2), $I(f) \geq \sum_{S\neq\emptyset}\hat{f}^2(S) = 1$.

Another application for example is to show that if f is boolean ,balanced and $I(f) = 1$ then f is linear, i.e. $f = \sum_i \hat{f}(i)\chi_i(x) = \sum_i \hat{f}(i)x_i$. By using (2), (4) and (5) we conclude that the only sets S for which $\hat{f}(S) \neq 0$, are the sets of size 1 which means that f is linear. Moreover since f is boolean then $\exists i$ such that $f(x) = x_i$ or $f(x) = -x_i$ (dictatorship) because otherwise if $\hat{f}(i), \hat{f}(j) \neq 0$ $(i \neq j)$ then for the possible four combinations of values for x_i, x_j we have four different values for the function, so the function cannot be boolean in contradiction.

Remark Next lecture we will show how to use this techniques to get robustness for this results.