CS 880: Pseudorandomness and Derandomization 2/13/2013

Lecture 7: Harmonic Analysis

Instructors: Holger Dell and Dieter van Melkebeek Scribe: Gautam Prakriya

In the previous lecture we defined and constructed small-bias generators and mentioned that they yield "almost" k-wise uniform generators. In this lecture we quantitatively investigate that connection using harmonic analysis of Boolean functions. We develop the basics of this important tool and illustrate its versatility with an application to list decoding for the Hadamard code, which complements our discussion from Lecture 5.

1 Basics of Harmonic Analysis

Classical setting. Harmonic analysis for real- or complex-valued functions on the domain of the reals, i.e., for functions $f : \mathbb{R} \to \mathbb{R}$ or $f : \mathbb{R} \to \mathbb{C}$, expresses f in a unique way as a sum, series, or integral of so-called harmonics. In the classical case, these harmonics are sine and cosine functions, or complex exponentials. The decomposition can be viewed as a basis transformation in the vector space of all functions f from the standard function basis to the basis of harmonics. The transformation is known as the Fourier transform, and has a number of useful properties, chief among them being the following.

1. It is orthogonal and – provided adequate scaling – even orthonormal, which means that it preserves angles and distances.

2. It transforms convolutions (defined in Exercise 2 below) into point-wise products, i.e., the Fourier transform of the convolution of two functions is just the point-wise product of the Fourier transforms.

Convolutions naturally appear in a number of contexts.

- In signal processing, convolutions occur when a filter is being applied to a signal. A filter is an operation that changes the signal; for example, a filter might smoothen the signal to dampen small rapid changes. The filter operation can be expressed as the convolution of the input signal with a function describing the filter.

  A straightforward computation of the convolution for signals of length n takes time $O(n^2)$. Alternately, one can first compute the Fourier transform of the operands, then multiply them point-wise, and finally apply the inverse Fourier transform to the result. Since the Fourier transform can be computed and inverted in time $O(n \log n)$ using the FFT algorithm, and point-wise products can be computed in time $O(n)$, this yields an $O(n \log n)$ algorithm. The improvement from $O(n^2)$ to $O(n \log n)$ combined with the ubiquity of convolutions in signal processing explains the importance of the Fourier transform there.

- A context that is more relevant for this course is that of probability theory, where the distribution of the sum of two independent random variables is given by the convolution of the individual distributions.

Generalization. Harmonic analysis with the properties 1 and 2 above has been generalized to real- or complex-valued functions on certain domains other than the reals. In particular, we consider domains that form a finite commutative group G, in which case the harmonics correspond to the characters of G.

Definition 1. *A* character *of a group* $(G, +)$ *is a homomorphism from* $(G, +)$ *to the group* $(\mathbb{C} \setminus \{0\}, \cdot)$ *of the complex numbers without zero under multiplication. In other words, a character of G is a mapping* $\chi : G \to \mathbb{C} \setminus \{0\}$ *such that*

$$(\forall x, y \in G) \quad \chi(x + y) = \chi(x) \cdot \chi(y)\,.$$

We consider the case in which G is a finite group. Then one can show in general that the characters form an orthonormal set with respect to the following inner product on the vector space of all functions $f, g : G \to \mathbb{C}$:

$$(f, g) \doteq \mathbb{E}_x\Big[f(x) \cdot \overline{g(x)}\Big]\,. \tag{1}$$

The expectation is with respect to the uniform distribution over G, and $\overline{x}$ denotes the complex conjugate of x.

Exercise 1. *Show that* (1) *defines a valid inner product, i.e., for every* $f, g, h : G \to \mathbb{C}$ *and* $\alpha, \beta \in \mathbb{C}$

- *Non-negativity:* (f, f) *is a nonnegative real, and* $(f, f) = 0$ *if and only if* f *is identically zero.*
- *Linearity:* $(\alpha f + \beta g, h) = \alpha(f, h) + \beta(g, h)$ *and* $(h, \alpha f + \beta g) = \overline{\alpha}(h, f) + \overline{\beta}(h, g)$*.*
- *Triangle inequality:* $\sqrt{(f + g, f + g)} \leq \sqrt{(f, f)} + \sqrt{(g, g)}$*.*

Note that the inner product (f, g) is a scaled version of the standard inner product $\langle f, g\rangle \doteq \sum_{x \in G} f(x)\overline{g(x)}$, namely $(f, g) = \frac{1}{|G|}\langle f, g\rangle$. This scaling makes the characters *orthonormal* rather than just *orthogonal*. In contrast, note that the canonical basis (consisting of all functions $x \mapsto I[x = a]$ for $a \in G$) is orthonormal with respect to the standard inner product $\langle\cdot, \cdot\rangle$, but only orthogonal with respect to $(\cdot, \cdot)$.

Thus, the set of characters of a finite group G forms an orthonormal set, that is, for all distinct characters χ and χ', we have $(\chi, \chi) = 1$ and $(\chi, \chi') = 0$. In particular, since functions that are orthogonal with respect to any valid inner product are linearly independent, we have that the characters form a linearly independent set in the vector space of all functions $f : G \to \mathbb{C}$. In the case of a finite commutative group G, the number of characters turns out to be equal to the number of elements in G. Since the dimension of the vector space of all functions $f : G \to \mathbb{C}$ equals $|G|$ (as can be seen from the canonical basis mentioned above), this implies that the characters form a basis. We can associate a character χ_a with each element $a \in G$. Then every function f can be expressed uniquely as a linear combination of the characters χ_a:

$$f = \sum_{a \in G} \hat{f}(a) \cdot \chi_a. \tag{2}$$

This expression is known as the *Fourier expansion* of f, the individual coefficients $\hat{f}(a)$ as the *Fourier coefficients* of f, and the function $\hat{f} : G \to \mathbb{C}$ as the *Fourier transform* of f. By taking the inner product with χ_a on both sides of (2), and using the orthonormality of the character basis, we obtain the following expression for the Fourier coefficients:

$$\hat{f}(a) = (f, \chi_a)\,.$$

By taking the inner product of the Fourier expansion of f with itself, and again using the orthonormality of the character basis, we obtain

$$(f, f) = \sum_{a \in G} |\hat{f}(a)|^2\,.$$

This equation is known as *Parseval's equality* and it expresses the preservation of distances under the Fourier transform.

We leave establishing the second chief property of the Fourier transform as an exercise.

Exercise 2. *For $f, g : G \to \mathbb{C}$, we define the convolution $f * g : G \to \mathbb{C}$ by*

$$(f * g)(x) = \mathbb{E}_y\Big[f(y) \cdot g(x - y)\Big]\,,$$

where the expectation is with respect to the uniform distribution over G. Show that $\widehat{(f * g)}(x) = \hat{f}(x) \cdot \hat{g}(x)$ holds for all $x \in G$.

Boolean setting. For our purposes, G is the group consisting of the Boolean cube $\{0,1\}^n = \mathbb{F}_2^n$ with component-wise addition. In this case, the characters are the parity functions rescaled from the range $\{0,1\}$ to the range $\{1,-1\}$. That is, for all $a \in G$,

$$\chi_a(x) = (-1)^{\langle a,x \rangle}$$

is a character of G, where $\langle a, x \rangle = \sum_i a_i x_i$ is the inner product over $\mathbb{F}_2$. The functions χ_a are characters because

$$\chi_a(x + y) = (-1)^{\langle a,x+y \rangle} = (-1)^{\langle a,x \rangle + \langle a,y \rangle} = (-1)^{\langle a,x \rangle} \cdot (-1)^{\langle a,y \rangle} = \chi_a(x) \cdot \chi_a(y).$$

We leave it as an exercise to check that there are no other characters. This also follows from the general fact that every commutative group has $|G|$ characters and the fact that there are $2^n = |G|$ distinct functions of the form χ_a.

We now explicitly prove that the functions χ_a form an orthonormal basis with respect to the inner product $(\cdot,\cdot)$. Since these functions only take on real values and we are only interested in real-valued functions, we can restrict the range to be $\mathbb{R}$ rather than $\mathbb{C}$.

Proposition 1. *The functions $\chi_a : \{0,1\}^n \to \mathbb{R}$ for $a \in \{0,1\}^n$ defined by $\chi_a(x) = (-1)^{\langle a,x \rangle}$ form an orthonormal basis for the vector space of all functions $f : \{0,1\}^n \to \mathbb{R}$ with respect to the inner product $(f, g) \doteq \mathbb{E}\big[f(x)g(x)\big]$, where the expectation is over the uniform distribution of $x \in \{0,1\}^n$.*

Proof. First note that $(\chi_a, \chi_b) = \mathbb{E}[(-1)^{\langle a,x \rangle}(-1)^{\langle b,x \rangle}] = \mathbb{E}[(-1)^{\langle a+b,x \rangle}]$. If $a \neq b$, then $a + b \neq 0$ and $(-1)^{\langle a+b,x \rangle}$ is 1 for half the values of x and 0 for the other half, which implies that $\mathbb{E}[(-1)^{\langle a+b,x \rangle}] = 0$. If $a = b$, then $a + b = 0$ and $\mathbb{E}[(-1)^{\langle a+b,x \rangle}] = 1$.

This shows that the functions χ_a form an orthonormal set and therefore a linearly independent set. To see that they form a basis, note that there are 2^n of them and the dimension of the vector space of all functions $f : \{0,1\}^n \to \mathbb{R}$ is 2^n. □

As a consequence of Proposition 1, we can write any function $f : \{0,1\}^n \to \mathbb{R}$ as (2), where the Fourier coefficients are given by

$$\hat{f}(a) = \mathbb{E}_{x \in_u \{0,1\}^n}\Big[f(x) \cdot \chi_a(x)\Big] = \mathbb{E}_{x \in_u \{0,1\}^n}\Big[f(x) \cdot (-1)^{\langle a,x \rangle}\Big],$$

and Parseval's equality tells us that

$$\mathbb{E}_{x \in_u \{0,1\}^n}\left[f(x)^2\right] = \sum_{a \in \{0,1\}^n} \hat{f}(a)^2 .$$

For a Boolean function $g : \{0,1\}^n \to \{0,1\}$, we typically consider the Fourier expansion of the rescaled variant $f : \{0,1\}^n \to \{-1,1\}$ given by $f(x) = (-1)^{g(x)}$. In that case, the Fourier coefficient

$$\hat{f}(a) = \Pr_{x \in_u \{0,1\}^n}\Big[g(x) = \langle a,x \rangle\Big] - \Pr_{x \in_u \{0,1\}^n}\Big[g(x) \neq \langle a,x \rangle\Big] \tag{3}$$

gives the correlation of $g(x)$ with the parity function with mask a, and Parseval's equality reads

$$\sum_{a \in \{0,1\}^n} \hat{f}(a)^2 = 1 . \tag{4}$$

This means that we can interpret $\hat{f}^2$ as a probability distribution over $\{0,1\}^n$.

2 List Decoding the Hadamard Code

We now present a simple application that illustrates the power of harmonic analysis for Boolean functions. Recall the list decoding problem for the Hadamard code: Given $\eta > 0$ and a received word g modeled as a function $g : \{0,1\}^k \to \{0,1\}$, we want to find the set S of all messages $a \in \{0,1\}^k$ whose codewords have a given degree of agreement with g. That is, S is the set of all a that satisfy

$$\Pr_{x \in_u \{0,1\}^n}\Big[g(x) = \langle a,x \rangle\Big] \geq \frac{1}{2} + \eta . \tag{5}$$

In Lecture 5, we developed a randomized algorithm for this problem, and as a byproduct of the analysis we concluded in Corollary 7 that the number $|S|$ of solutions a is $O(k/\eta^2)$. We now show how to improve this upper bound to one that is independent of k, namely $|S| \leq \frac{1}{4\eta^2}$.

For this, consider $f(x) = (-1)^{g(x)}$. By (3), the requirement (5) is equivalent to $\hat{f}(a) \geq 2\eta$. By Parseval's equality (4),

$$|S| \cdot (2\eta)^2 \leq \sum_{a \in S} \hat{f}(a)^2 \leq \sum_{a \in \{0,1\}^n} \hat{f}(a)^2 = 1,$$

which implies that $|S| \leq \frac{1}{4\eta^2}$.

3 Almost k-Wise Uniformity

We now use harmonic analysis to quantitatively show that every small-bias distribution is "almost" k-wise uniform. Recall that for a distribution D on $\{0,1\}^r$ and $a \in \{0,1\}^r$, we defined

$$\text{bias}_a(D) \doteq \mathbb{E}\left[(-1)^{\langle a,D \rangle}\right] = \Pr\left[\langle a,D \rangle = 0\right] - \Pr\left[\langle a,D \rangle = 1\right],$$

and D is β-biased if $\left|\mathrm{bias}_a(D)\right| \leq \beta$ holds for every nonzero a. In the previous lecture, we showed that D is β-biased if and only if D is ϵ-pseudorandom for polynomials of degree 1 with $\epsilon = \beta/2$. We also argued that, if a distribution D on $\{0,1\}^r$ is ϵ-pseudorandom for polynomials of degree k, then for every subset $I \subseteq [r]$ with $|I| = k$,

(i) $(\forall y \in \{0,1\}^k) \quad \Big|\Pr[D|_I = y] - \Pr[U_k = y]\Big| \leq \epsilon$, and

(ii) $d_{\mathrm{stat}}(D|_I, U_k) \leq 2^k \epsilon$.

We now show the following strengthening.

Proposition 2. *Let D be a distribution on $\{0,1\}^r$ such that $|\mathrm{bias}_a(D)| \leq \beta$ for every nonzero $a \in \{0,1\}^r$ of weight at most k. Then for every $I \subseteq [r]$ with $|I| = k$, we have*

(i) $(\forall y \in \{0,1\}^k) \quad \Big|\Pr[D|_I = y] - \Pr[U_k = y]\Big| \leq \left(1 - \frac{1}{2^k}\right) \cdot \beta$, *and*

(ii) $d_{\mathrm{stat}}(D|_I, U_k) \leq \sqrt{2^k - 1} \cdot \frac{\beta}{2}$.

Note that Proposition 2 strengthens the statements from last lecture in several ways. It shows that requiring ϵ-pseudorandomness for polynomials of degree k is an overkill; degree 1 suffices. Moreover, it quantitatively improves the bound for (ii).

Proof (of Proposition 2). Let $f : \{0,1\}^k \to [0,1]$ be defined by $f(y) = \Pr[D|_I = y]$, and consider the Fourier expansion $f = \sum_{b \in \{0,1\}^k} \hat{f}(b) \cdot \chi_b$. We have that

$$\begin{aligned}
\hat{f}(b) &= \mathbb{E}_{y \in_u \{0,1\}^k}[f(y)(-1)^{\langle b,y\rangle}] \\
&= \frac{1}{2^k} \cdot \sum_{y \in \{0,1\}^k} f(y)(-1)^{\langle b,y\rangle} \\
&= \frac{1}{2^k} \cdot \mathbb{E}\left[(-1)^{\langle b, D|_I\rangle}\right] \\
&= \frac{\mathrm{bias}_b(D|_I)}{2^k} \\
&= \frac{\mathrm{bias}_a(D)}{2^k},
\end{aligned}$$

where $a \in \{0,1\}^r$ is such that $a|_I = b$ and $a|_{\overline{I}} = 0$. Observe that a has weight at most k, and that $a \neq 0$ if and only if $b \neq 0$. Therefore, we can conclude that $\hat{f}(0) = \frac{\mathrm{bias}_0(D)}{2^k} = \frac{1}{2^k}$, and by our hypothesis that $|\hat{f}(b)| \leq \beta/2^k$ for every $b \neq 0$. By considering the uniform distribution $D = \frac{1}{2^k}$, the above derivation also shows that $\widehat{\frac{1}{2^k}}(b)$ equals $\frac{1}{2^k}$ for $b = 0$, and vanishes for every nonzero b. Thus, we also have that

$$\left|\widehat{\left(f - \frac{1}{2^k}\right)}(b)\right| \qquad \begin{cases} = 0 & \text{if } b = 0, \\ \leq \frac{\beta}{2^k} & \text{if } b \neq 0. \end{cases} \tag{6}$$

For any $y \in \{0,1\}^k$, we have that

$$\begin{aligned}
\Big|\Pr[D|_I = y] - \Pr[U_k = y]\Big| &= \left|f(y) - \frac{1}{2^k}\right| \\
&= \left|\sum_{0 \neq b \in \{0,1\}^k} \hat{f}(b)\chi_b(y)\right| \\
&\leq \sum_{0 \neq b \in \{0,1\}^k} |\hat{f}(b)\chi_b(y)| \\
&= \sum_{0 \neq b \in \{0,1\}^k} |\hat{f}(b)| \\
&\leq (2^k - 1) \cdot \frac{\beta}{2^k} \\
&= (1 - \frac{1}{2^k}) \cdot \beta,
\end{aligned}$$

which establishes part (i). For part (ii) we have that

$$\begin{aligned}
d_{\text{stat}}(D|_I, U_k) &= \frac{1}{2} \sum_{y \in \{0,1\}^k} |f(y) - \frac{1}{2^k}| \\
&= 2^{k-1} \cdot \mathbb{E}_{y \in_u \{0,1\}^k}[|f(y) - \frac{1}{2^k}|] \\
&\leq 2^{k-1} \cdot \sqrt{\mathbb{E}_{y \in_u \{0,1\}^k}[(f(y) - \frac{1}{2^k})^2]} && \text{(by Cauchy–Schwarz)} \\
&= 2^{k-1} \cdot \sqrt{(f - \frac{1}{2^k}, f - \frac{1}{2^k})} && \text{(by the definition of the inner product)} \\
&= 2^{k-1} \cdot \sqrt{\sum_{b \in \{0,1\}^k} (\widehat{f - \frac{1}{2^k}}(b))^2} && \text{(by Parseval's equality)} \\
&\leq 2^{k-1} \cdot \sqrt{\sum_{0 \neq b \in \{0,1\}^k} (\frac{\beta}{2^k})^2} && \text{(by (6))} \\
&= \sqrt{2^k - 1} \cdot \frac{\beta}{2}. && \square
\end{aligned}$$

We quantify the notion of "almost" k-wise uniformity as follows.

Definition 2 (Almost k-wise uniformity). *A distribution D on $\{0,1\}^r$ is ϵ-pseudorandom for k-wise uniformity if, for every $I \subseteq [r]$ with $|I| = k$, we have $d_{\text{stat}}(D|_I, U_k) \leq \epsilon$.*

Recall that Theorem 4 from Lecture 6 gives an efficient β-bias generator with seed length $O(\log(r/\beta))$. Applying part (ii) of Proposition 2 to this construction yields the following result.

Theorem 3. *For all integers r and $k \in [r]$, and every real $\epsilon > 0$, there exists an efficiently computable ϵ-pseudorandom generator $G : \{0,1\}^\ell \to \{0,1\}^r$ for k-wise uniformity with $\ell = O(k + \log(1/\epsilon) + \log(r))$.*

Note that the k term in the seed length of Theorem 3 is additive, as opposed to the seed length $O(k \cdot \log(r))$ in our construction of perfectly k-wise uniform generators from Lecture 4. In the next lecture, we will see how we can again use small-bias generators to further reduce the seed length of ϵ-PRGs for k-wise uniformity from $O(k + \log(1/\epsilon) + \log(r))$ in Theorem 3 to $O(k + \log(1/\epsilon) + \log\log(r))$.