0368-4159: First Course in Derandomization 17/12/18 – Lecture 10

Small Bias

Lecturer: Amnon Ta-Shma

Scribe: Dean Doron

1 The Fourier transform

Given a finite set A, the set $\mathbb{C}^A = \{f : A \to \mathbb{C}\}$ is a vector space of dimension $|A|$ over $\mathbb{C}$. There is a "natural" basis for this vector space. In this lecture we present another, very useful, basis for this vector space when A is a *commutative group*. We then focus on the case when $A = \mathbb{Z}_2^n$, i.e., on representations of boolean functions.

1.1 Over general domains

Let G be a finite group with operation $+$ and identity 0. Then, the group algebra $\mathbb{C}[G]$ is the set of all functions $f \in \mathbb{C}[G]$. Obviously, it is a vector space on G over the field $\mathbb{C}$ of dimension $|G|$ and a natural basis for $\mathbb{C}[G]$ is

$$\mathbf{1}_g(x) \;=\; \begin{cases} 1, & x = g \\ 0, & \text{otherwise} \end{cases}$$

for every $g \in G$. It is also an inner-product space under the inner product

$$\langle f_1, f_2 \rangle = \mathop{\mathbb{E}}_{x \in G} \overline{f_1(x)} f_2(x) = \frac{1}{|G|} \sum_{x \in G} \overline{f_1(x)} f_2(x),$$

and it is easy to see that the basis $\{\mathbf{1}_g\}_{g \in G}$ is an orthogonal basis under this inner product. Often, one writes g instead of $\mathbf{1}_g$ and in this notation an element $f \in \mathbb{C}[G]$ is represented as $\sum_{g \in G} a_g g$ (which is often the notation used in quantum computation).

We now introduce another basis, that contains only functions that are homomorphisms from G to $\mathbb{C}^\times$.

Definition 1. *A* character *of the finite group G is a homomorphism $\chi : G \to \mathbb{C}^\times$, i.e., $\chi(x+y) = \chi(x)\chi(y)$ for every $x, y \in G$, where the addition is the group operation in G, and the multiplication is the group operation in $\mathbb{C}^\times$.*

We have the following easy facts:

Claim 2. *Let G be a finite group. Then:*

1. *$\chi_{\text{trivial}}(x) = 1$ for every $x \in G$ is a character. It is called the trivial character.*
2. *If χ_1 and χ_2 are characters of G then so is $\chi_1 \cdot \chi_2$ (where $\chi_1 \cdot \chi_2(x) = \chi_1(x)\chi_2(x)$).*
3. *For every character χ of G and $x \in G$, $|\chi(x)| = 1$ (the absolute norm is of course in $\mathbb{C}$). In particular, if we define $\overline{\chi_1}(x) = \overline{\chi_1(x)}$, then $\overline{\chi_1}$ is also a character and $\chi \cdot \overline{\chi} = \chi_{\text{trivial}}$.*

4. *This implies that* $\widehat{G} = \{\chi \in \mathbb{C}[G] \mid \chi \text{ is a character}\}$ *is an Abelian group, with identity as in item (1), multiplication as in item (2) and inverse as in item (3).*

5. *Let* χ *be a non-trivial character. Then,* $\mathbb{E}[\chi] = 0$*. This means that every non-trivial character is orthogonal to the trivial character.*

We can then show:

Claim 3. *Let* G *be a finite group. The set of all characters of* G *is orthonormal.*

Proof. First, note that $\langle \chi, \chi \rangle = \mathbb{E}[|\chi|^2] = 1$. Next, take χ_1 and χ_2 be two distinct characters of G. Then, $\langle \chi_1, \chi_2 \rangle = \mathbb{E}[\overline{\chi_1}\chi_2]$. However, $\overline{\chi_1}\chi_2$ is itself a character, and $\overline{\chi_1}\chi_2 = \chi_1^{-1}\chi_2 \not\equiv 1$ since they are distinct. Thus, $\mathbb{E}[\overline{\chi_1}\chi_2] = 0$. □

As a consequence, G has at most $\dim(\mathbb{C}[G]) = |G|$ characters.

We will soon see that when G is Abelian, $\widehat{G}$ has a full set of characters. The resulting orthonormal basis for $G[\mathbb{C}]$ is called the *Fourier basis*, and the linear transformation between the natural basis and the Fourier basis is called the *Fourier transform*. Thus, every $f \in G[\mathbb{C}]$ can be (uniquely) written as $f = \sum_S \hat{f}_S \cdot \chi_S$, where χ_S run over all characters in $\widehat{G}$. The coefficients $\hat{f}_S$ are called the Fourier coefficients of f.

Let us see some examples.

For $G = \mathbb{Z}_2$, it is easy to check that $\chi_1 \equiv 1$ and $\chi_2(x) = (-1)^x$ are characters (and we know that there are no more than 2). Next consider $G = \mathbb{Z}_m$ with addition modulo m. If χ is a character, and $x \in G$, then $\chi(x)^m = \chi(mx) = \chi(0) = 1$, hence, $\chi(x)$ is an m-th root of unity. Denote $\omega = e^{\frac{2\pi i}{m}}$. For $0 \le j < m$, define $\chi_j : \mathbb{Z}_m \to \mathbb{C}$ by $\chi_j(x) = \omega^{jx}$. It is easy to see that these are distinct characters of $\mathbb{Z}_m$ and since we have m of them, they are all the characters and $|\widehat{G}| = |G|$.

Let $f : \mathbb{Z}_m \to \mathbb{C}$. By now, we know that its Fourier expansion is given by $f(n) = \sum_{k=0}^{m-1} \hat{f}_k \omega^{kn}$. If we treat f and $\hat{f}$ as vectors in $\mathbb{C}^m$, we get

$$f = \begin{pmatrix} \omega^{0\cdot 0} & \omega^{0\cdot 1} & \cdots & \omega^{0\cdot(m-1)} \\ \omega^{1\cdot 0} & \omega^{1\cdot 1} & \cdots & \omega^{1\cdot(m-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \omega^{(m-1)\cdot 0} & \omega^{(m-1)\cdot 1} & \cdots & \omega^{(m-1)\cdot(m-1)} \end{pmatrix} \cdot \hat{f},$$

and the above matrix is called the Fourier matrix.

We now consider group products. Say $(A, \cdot)$, $(B, \cdot)$ are two groups. A and B are not necessarily Abelian, and we denote their operation by $\cdot$ rather than $+$ to (somewhat) distinguish them from the Abelian case. Let $G = A \times B$ (i.e., the group operation in G is $(a_1, b_1) \cdot (a_2, b_2) = (a_1 \cdot a_2, b_2 \cdot b_2)$). For $f \in \mathbb{C}[A]$ and $g \in \mathbb{C}[B]$, define $f \otimes g \in \mathbb{C}[A \times B]$ by $(f \otimes g)(a, b) = f(a)g(b)$. Then:

Claim 4. *If* $f \in \widehat{A}$ *and* $g \in \widehat{B}$ *then* $f \otimes g \in \widehat{A \times B}$*. Also, all pairs* $f_i \otimes g_j$ *for* $f_i \in \widehat{A}$ *and* $g_j \in \widehat{B}$ *are distinct.*

Back to the Abelian case, we see that if A and B are finite Abelian groups than Claim 4 gives us $|A| \cdot |B| = |A \times B|$ characters, and so we have a full set of characters. As every Abelian group can be decomposed as a product of cyclic groups, we have:

Exercise 5. *Let G be a finite Abelian group. Then $G \simeq \widehat{G}$.*

So what are the characters of $G = \mathbb{Z}_2^n = \mathbb{Z}_2 \times \ldots \times \mathbb{Z}_2$? By the above discussion, for every $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_2^n$ we have the character

$$\chi_\alpha(x) = (\chi_{\alpha_1} \otimes \ldots \otimes \chi_{\alpha_n})(x_1, \ldots, x_n) = \prod_i \chi_{\alpha_i}(x_i) = \prod_i (-1)^{\alpha_i x_i} = (-1)^{\sum_i x_i \alpha_i}.$$

Equivalently we could say that for every $S \subseteq [n]$ there is the character $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. The trivial character is the character of the empty set. Parity is the character of the full set (more precisely, $(-1)^{\text{parity}}$) and every function $f : \{0,1\}^n \to \mathbb{C}$ can be written as

$$f(a) \;=\; \sum_{S \subseteq [n]} \hat{f}_S \chi_S(a) \;=\; \sum_{S \subseteq [n]} \hat{f}_S \cdot (-1)^{\langle \mathbf{1}_S, a \rangle}.$$

We also see that the linear transformation converting the natural basis to the Fourier basis or vice versa, is the Hadamard matrix.

1.2 Some useful properties of the Fourier transform

Theorem 6. *For any $f, g : G \to \mathbb{C}$,*

- $\hat{f}_S = \langle \chi_S, f \rangle$.
- *(Parseval's Theorem)* $\langle f, f \rangle = \sum_S |\hat{f}_S|^2$.
- *(Plancherel's Theorem)* $\langle f, g \rangle = \sum_S \overline{\widehat{f}_S} \widehat{g}_S$.

For example, let us take $f : \{0,1\}^3 \to \{0,1\}$ to be the majority function. Verify that $\hat{f}(\emptyset) = \frac{1}{2}$, $\hat{f}(\{1\}) = \hat{f}(\{2\}) = \hat{f}(\{3\}) = -\frac{1}{4}$, $\hat{f}(\{1,2\}) = \hat{f}(\{1,3\}) = \hat{f}(\{2,3\}) = 0$ and $\hat{f}(\{1,2,3\}) = \frac{1}{4}$. Also, you can check that Parseval's theorem holds, as $\langle f, f \rangle = \frac{1}{2}$.

Another useful property (used, e.g., in the fast FFT algorithm for multiplying two polynomials) is that convulsion in the standard basis is transformed to coordinate-wise product in the Fourier basis.

Definition 7. *Let $f, g : G \to \mathbb{C}$. Define $F * g : G \to \mathbb{C}$ by*

$$f * g(a) = \mathop{\mathbb{E}}_{\substack{(x,y) \in G \times G: \\ x+y=a}} f(x) g(y) = \mathop{\mathbb{E}}_{x \in G} f(x) g(a-x).$$

Claim 8. $\widehat{f * g}_S = \widehat{f}_S \cdot \widehat{g}_S$.

Proof.

$$\begin{aligned}
\widehat{f * g}_S &= \mathop{\mathbb{E}}_{x \in G} \overline{\chi_S(x)} f * g(x) = \mathop{\mathbb{E}}_{x \in G} \mathop{\mathbb{E}}_{y \in G} \overline{\chi_S(x)} f(y) g(x-y) \\
&= \mathop{\mathbb{E}}_{z \in G} \mathop{\mathbb{E}}_{y \in G} \overline{\chi_S(z+y)} f(y) g(z) = \mathop{\mathbb{E}}_{z \in G} \mathop{\mathbb{E}}_{y \in G} \overline{\chi_S(z)} \cdot \overline{\chi_S(y)} \cdot f(y) \cdot g(z) \\
&= \mathop{\mathbb{E}}_{z \in G} \overline{\chi_S(z)} g(z) \mathop{\mathbb{E}}_{y \in G} \overline{\chi_S(y)} f(y) = \widehat{g}_S \cdot \widehat{f}_S.
\end{aligned}$$

where the second line is by the change of variable $z = x - y$. □

We next give an intuitive explanation on the *meaning* of these numbers.

1.3 Some examples

We now consider several $f : \{0,1\}^n \to \{0,1\}$.

- For the parity function, $f(x_1, \ldots, x_n) = \sum x_i \bmod 2$. We have $\hat{f}_\emptyset = \frac{1}{2}$ and $\hat{f}_{[n]} = -\frac{1}{2}$.
- For the and function, $f(x_1, \ldots, x_n) = \wedge_{i=1}^n x_i$ we have $f(x) = \Pi_{i=1}^n x_i = \Pi_{i=1}^n (\frac{1-\chi_i}{2}) = \sum_{S \subseteq [n]} 2^{-n}(-1)^{|S|}\chi_S$.

2 ε-biased sets

A set $T \subseteq \Lambda$ ε-fools a function $f : \Lambda \to \{0,1\}$ if $|\Pr_{x\in\Lambda}[f(x) = 1] - \Pr_{x\in T}[f(x) = 1]| \leq \varepsilon$. A set T ε-fools a class of functions $\mathcal{C}$ if it ε-fools every $f \in \mathcal{C}$. A set $T \subseteq \{0,1\}^k$ is called ε-biased if it ε-fools all functions $f : \{0,1\}^k \to \{0,1\}$ that are linear over $\mathbb{F}_2$. Formally:

Definition 9. *Let $T \subseteq \{0,1\}^k$. For a nonzero $w \in \{0,1\}^k$ we denote*

$$\mathrm{Bias_w(T)} \;=\; \left|\Pr_{\mathrm{s\in T}}[\langle \mathrm{w,s}\rangle = 0] - \Pr_{\mathrm{s\in T}}[\langle \mathrm{w,s}\rangle = 1]\right| = 2\left|\frac{1}{2} - \Pr_{\mathrm{s\in T}}[\langle \mathrm{w,s}\rangle = 1]\right|.$$

The bias of T is $\mathrm{Bias(T)} = \max_{\mathrm{w}\neq 0} \mathrm{bias_w(T)}$. *We say that T is ε-biased if* $\mathrm{Bias(T)} \leq \varepsilon$.

An ε-biased set tries to fool a class of functions using samples from a small set (in other words, we try to achieve pseudo-randomness with respect to a (very) limited class of tests). It is then natural to ask how *small* can ε-biased sets be. We shall soon answer this. But first, we interpret ε-bias using Fourier representation.

2.1 ε bias and the Fourier transform

Let X be a distribution over $\{0,1\}^n$ and $w \in \{0,1\}^n$. Then,

$$\begin{aligned}\mathrm{Bias_w(X)} \;&=\; \left|\Pr_{s\sim X}[\langle w,s\rangle = 0] - \Pr_{s\sim X}[\langle w,s\rangle = 1]\right| \;=\; \left|\sum_{s\in\{0,1\}^n} (-1)^{\langle s,w\rangle}\cdot \Pr[X=s]\right| \\ &=\; \left|\sum_{s\in\{0,1\}^n} X(s)\chi_w(s)\right| \;=\; 2^n\cdot|\langle X,\chi_w\rangle| \;=\; 2^n\cdot|\hat{X}_w|.\end{aligned}$$

Thus, we can redefine ε bias in the Fourier language.

Definition 10. *(equivalent to Def 9) Let $T \subseteq \{0,1\}^k$ and X the flat distribution over S. We say that T is ε-biased if $|\hat{X}_S| \leq \varepsilon 2^{-n}$ for all $S \neq \emptyset$.*

We prove:

Theorem 11 ([5, 4])**.** *Let X be distribution over $\{0,1\}^n$. Then $\|X - U_n\|_2 \leq \mathrm{Bias(X)}$ and $\|X - U_n\|_1 \leq 2^{n/2}\cdot \mathrm{Bias(X)}$.*

Proof. Express $X = \sum_w \hat{X}_w \chi_w$. Now, $X - U_n = \sum_{w \neq \emptyset} \hat{X}_w \chi_w$ (Why?). We bound the ℓ_2 norm of $X - U_n$. We have:

$$\|X - U_n\|_2^2 = 2^n \langle X - U_n, X - U_n \rangle = 2^n \sum_{w \neq \emptyset} |\hat{X}_w|^2 \leq 2^n 2^n (\varepsilon 2^{-n})^2 = \varepsilon^2.$$

The bound on the ℓ_1 norm follows from Cauchy-Schwartz. □

In particular we see that if X has zero bias than X must be the uniform distribution.

3 ε bias and binary error correcting codes

Definition 12. *An $[n, k]$ error correcting code C is ε-balanced if the Hamming weight of every non-zero codeword in C is between $\left(\frac{1}{2} - \varepsilon\right) n$ and $\left(\frac{1}{2} + \varepsilon\right) n$.*

Claim 13. *$M_{n \times k}$ is a generator matrix of an $[n, k]_2$ error correcting code that is ε-balanced, iff $\{r_i \mid r_i$ is the i'th row of $M\} \subseteq \{0,1\}^k$ is ε-biased.*

Proof. Let M be a generator matrix of an $[n, k]$ ε-balanced code C. For every $x \in \{0,1\}^k$, Mx contains at least $\left(\frac{1}{2} - \varepsilon\right) n$ nonzero entries and at most $\left(\frac{1}{2} + \varepsilon\right) n$. Hence, if we choose a row M_r of M uniformly at random, $\Pr_{r \in [n]}[\langle x, M_r \rangle = 1] \in [\frac{1}{2} \pm \varepsilon]$. It is then clear that the rows of M constitutes an ε-biased set in $\{0,1\}^k$ of size n. The other direction holds as well. We leave this to the reader. □

We are now ready to prove non-explicit existence.

Claim 14. *For every k, there exists an ε-biased set $T \subseteq \{0,1\}^k$ of size $n = O(\frac{k}{\varepsilon^2})$.*

Proof. Choose the entries of A, a binary matrix of dimension $n \times k$, uniformly at random. Fix a nonzero $x \in \{0,1\}^k$ and let W_x be the Hamming weight of Ax. That is, $W_x = \sum_{i=1}^n \langle A_i, x \rangle$ where the inner-product is modulo 2.

For a fixed non-zero x, $\mathbb{E}[W_x] = \frac{n}{2}$ (why?). By Chernoff, the probability that Ax is bad is at most

$$\Pr\left[\left|\frac{1}{n} W_x - \frac{1}{2}\right| \geq \varepsilon\right] \;\leq\; 2e^{-2n\varepsilon^2}.$$

By the union bound, the probability that A is a generator matrix for an unbalanced code is at most $2^k \cdot 2e^{-2n\varepsilon^2} \leq 2^{k+1-2n\varepsilon^2} < 1$, for $n \geq \frac{k}{\varepsilon^2}$. □

Non-explicitly the lower bound is $n = \Omega\left(\frac{k}{\epsilon^2 \log(\frac{1}{\varepsilon})}\right)$, and the same lower bound holds for $[n, k, \frac{1}{2} - \epsilon]_2$ codes (that are not necessarily ε-balanced, i.e., they may have high weight codewords).

3.1 Explicit ε-bias constructions

In the last lecture we saw an explicit construction with about size $O(\frac{k}{\epsilon^4})$ (when seeing the error correcting code amplifications). Next, we show a construction that achieves the incomparable bound $n = O(\frac{k^2}{\varepsilon^2})$, due to Alon et al. [1]. As we said before the best non-explicit bound is $O(\frac{k}{\epsilon^2})$.

The AGHP [1] construction is Reed-Solomon concatenated with Hadamard. Specifically, we have the following ingredients:

- The outer code: An $R = [q = \frac{k_1}{\varepsilon}, k_1, 1-\varepsilon]_q$ Reed-Solomon code, for q that is a power of 2.
- The inner code: An $H = [q, \log(q), \frac{1}{2}]_2$ Hadamard code.

Our code is the concatenation of the two codes, namely,

$$H(R(x)_1), \ldots, H(R(x)_q).$$

Then, the concatenated code $R \circ H$ is a $[n = q^2, k = k_1 \log q, \frac{1}{2}(1-\varepsilon)]$ linear error correcting code. Now, $q = \frac{k_1}{\varepsilon} = \frac{k}{\varepsilon \log q}$ and so $n \leq (\frac{k}{\varepsilon})^2$. This shows that in the code there are no nonzero codewords of weight smaller than $\frac{1}{2}(1-\varepsilon)$. In fact, the concatenated code also does not have any codewords of length more than $\frac{1}{2}$ (why?) and so we get an ε-balanced code as needed.

4 k-wise and almost k-wise independence

Definition 15. *Let X be a distribution over $\{0,1\}^n$.*

- *We say X is* (k, ε)-biased*, if it is at most ε-biased with respect to all non-empty, linear tests of size at most k.*
- *We say X is* (k, ε)-wise independent *if for all $S \subseteq [n]$ of size k, $|X_{|S} - U_k|_1 \leq \varepsilon$.*

Equivalently, X is (k, ε)-*biased*, iff $|\hat{X}_S| \leq \varepsilon 2^{-n}$ for all S s.t. $1 \leq |S| \leq k$. X is k-wise independent iff $|\hat{X}_S| = 0$ for all S s.t. $1 \leq |S| \leq k$.

Theorem 16 ([5])**.** *There exists an distribution that is (k, ε)-biased over $\{0,1\}^n$ and has support size at most $O\left(\frac{k \log n}{\varepsilon^2}\right)$. There are explicit constructions with $O\left(\frac{(k \log n)^2}{\varepsilon^2}\right)$ or $O\left(\frac{k \log n}{\varepsilon^4}\right)$.*

Proof. For the construction we combine two ingredients that we already have: k-wise independence and ε-bias. Let

- A of size $n \times h$ be the generator matrix of a k-wise sample space. We saw how to construct A with $h = k \log n$ (and in fact, over $\mathbb{F}_2$ we can even get $h = \frac{1}{2} k \log n$).
- Sample $b \in B$, where $B \subseteq \{0,1\}^h$ is an ε-biased sample space. We saw how to construct B with support size $\left(\frac{h}{\varepsilon}\right)^2$ or even $O\left(\frac{h}{\varepsilon^4}\right)$ (non explicitly, $O\left(\frac{h}{\varepsilon^2}\right)$).

The construction: Sample $b \in B$ output $Ab \in \{0,1\}^n$.

Let $S \subseteq [n]$ be a set of size at most k. We want to bound $\mathrm{Bias}_S(\mathrm{Ab})$. Let A_i be the i-th row of A. It holds that:

$$\oplus_{i \in S} \ (Ab)_i = \oplus_{i \in S} \ \langle A_i, b \rangle = \left\langle \sum_{i \in S} A_i, b \right\rangle .$$

The vectors $\{A_i\}_{i \in S}$ are linearly independent and so $\sum_{i \in S} A_i$ is a nonzero test. As B is an ε-biased distribution, $\Pr_{b \in B}[\oplus_{i \in S} Ab_i = 1] \in [\frac{1}{2} \pm \varepsilon]$. The support size when $|B| = \frac{h^{c_1}}{\varepsilon^{c_2}} = O(\frac{(k \log n)^{c_1}}{\varepsilon^{c_2}})$. □

It therefore follows:

Corollary 17. *There exists an explicit distribution that is (k, ε)-wise independent over $\{0,1\}^n$ and has support size at most $2^k \left(\frac{k \log n}{\varepsilon}\right)^2$.*

Proof. By Theorem 11, an (k, ε')-biased distribution is $(k, \varepsilon' \cdot 2^{k/2})$-wise independent. Setting $\varepsilon' = 2^{-k/2}\varepsilon$, we are finished. □

5 The Fourier transform as a multilinear representation

We now choose to work with the group $\mathbb{Z}_{2,\cdot} = (\{1,-1\}, \cdot)$ instead of $\mathbb{Z}_{2,+} = (\{0,1\}, + \bmod 2)$ as we did so far. The two groups are isomorphic with the isomorphism $\psi : b \mapsto (-1)^b$. The two characters of $\mathbb{Z}_{2,\cdot}$ are $\mathbf{1}(x) = 1$ and $\mathbf{x}(x) = x$. Consequently, the characters of $\mathbb{Z}_{2,\cdot}^n$ are $\prod_{i \in S} x_i$. If we take $g : \mathbb{Z}_{2,\cdot}^n \to \mathbb{C}$, then its Fourier representation tells us how to open g as a multi-linear function over $\mathbb{C}$.

We identify a boolean function $f : \{0,1\}^n \to \{0,1\}$ with a function $g : \{-1,1\}^n \to \{1,-1\}$ defined by

$$g(\psi(b_1), \ldots, \psi(b_n)) = \psi(f(b_1, \ldots, b_n)).$$

It turns out that the Fourier representation of f in $\mathbb{Z}_{2,+}^n$ is closely related to the Fourier representation of g in $\mathbb{Z}_{2,\cdot}^n$:

Exercise 18. *Suppose $f(x) = \sum_S \hat{f}_S \chi_S(x)$ and $g(y) = \sum_S \hat{g}_S \prod_{i \in S} y_i$. Then $\hat{g}_\emptyset = 1 - 2\hat{f}_\emptyset$ and $\hat{g}_S = -2\hat{f}_S$ for all $S \neq \emptyset$.*

Hint: $\psi(b) = 1 - 2b$.

Thus, the Fourier expansion of f tells how g is represented as a multilinear function. The translation between f and g is linear ($f = \frac{1}{2}(1-g)$), as is the translation between the variables ($y_i = 1 - 2x_i$). In particular, $\max_{S : \hat{f}(S) \neq 0} |S|$ is the degree of the multilinear polynomial computing g over $\mathbb{C}$. For example, the Parity function $f(x_1, \ldots, x_n) = \sum_i x_i$ is linear over $\mathbb{F}_2$ but has degree n over $\mathbb{C}$.

From this discussion it is clear that the Fourier representation can help determine how close a function on $\{0,1\}^n$ is to being linear, or to a low-degree multilinear function over $\mathbb{C}$.

6 Linearity testing

Let L be the set of linear functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$. The BLR linearity test [3] does the following:

Input : A function $g : \mathbb{F}_2^n \to \mathbb{F}_2$ given by a black-box access, i.e., on query $y \in \mathbb{F}_2$ we get as answer $g(y) \in \mathbb{F}_2$.

Algorithm : Pick $x, y \in \mathbb{F}_2^n$ uniformly at random. Query $g(x), g(y), g(x+y)$.

Output : Output "yes" if $g(x+y) = g(x) + g(y)$ and "no" otherwise.

Clearly, uf $g \in L$ the algorithm always answers "yes". The surprising thing is the following "robustness" of the test:

Lemma 19. $\Pr[\text{"}no\text{"}] \geq dist(g, L)$.

Proof. Let $g : \mathbb{F}_2^n \to \{-1, 1\}$ be defined by $g(x) = (-1)^{f(x)}$. Then,

$$\begin{aligned}\Pr[\text{"}yes\text{"}] - \Pr[\text{"}no\text{"}] &= \Pr_{x,y\in\mathbb{F}_2^n}[f(x+y) = f(x) + f(y)] - \Pr_{x,y\in\mathbb{F}_2^n}[f(x+y) \neq f(x) + f(y)] \\ &= \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_2^n} g(x)g(y)g(x+y).\end{aligned}$$

Now,

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_2^n} g(x)g(y)g(x+y) &= \mathop{\mathbb{E}}_x g(x) \mathop{\mathbb{E}}_y g(y)g(x+y) = \mathop{\mathbb{E}}_x g(x) \cdot g * g(x) = \langle g, g * g\rangle \\ &= \sum_{S\subseteq[n]} \widehat{g}_S \cdot \widehat{g * g}_S = \sum_{S\subseteq[n]} \widehat{g}_S^3 = \sum_{S\subseteq[n]} \widehat{g}_S \cdot \widehat{g}_S^2\end{aligned}$$

For example, if f is linear, the g is a character, say, $g = \chi_T$ and then the expectation is indeed $1^3 = 1$. For a general g:

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_2^n} g(x)g(y)g(x+y) &= \sum_{S\subseteq[n]} \widehat{g}_S^3 \leq \max_T \widehat{g}_T \sum_{S\subseteq[n]} \widehat{g}_S^2 \\ &= \max_T \widehat{g}_T \cdot \langle g, g\rangle = \max_T \widehat{g}_T.\end{aligned}$$

However,

$$\begin{aligned}\hat{g}_T &= \langle \chi_T, g\rangle = \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n} (-1)^{\langle T,x\rangle} g(x) = \mathop{\mathbb{E}}_{x\in\mathbb{F}_2^n} (-1)^{\langle T,x\rangle} (-1)^{f(x)} \\ &= \Pr_{x\in\mathbb{F}_2^n}[f(x) = \langle T, x\rangle] - \Pr_{x\in\mathbb{F}_2^n}[f(x) \neq \langle T, x\rangle] = 1 - 2dist(f, \langle T, \cdot\rangle).\end{aligned}$$

Hence, $\max_T \hat{g}_T = 1 - 2dist(f, L)$. Hence,

$$1 - 2\Pr[\text{"}no\text{"}] = \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_2^n} g(x)g(y)g(x+y) \leq 1 - 2dist(f, L),$$

and $\Pr[\text{"}no\text{"}] \geq dist(f, L)$. □

For more reading see Section 1.6 of Ryan O'Donnell's book [6] (also [2]).

7 ε-bias and Cayley graphs over Abeliean groups

Lemma 20. *Let G be a group of order n. Suppose M is an $n \times n$ matrix such that $M[x,y] = f(xy^{-1})$. Then every character $\chi \in \hat{G}$ is an eigenvector of M with eigenvalue $|G| \cdot \hat{f}_\chi = |G|\langle \chi, f\rangle$.*

Proof.

$$\begin{aligned}(M\chi)_x &= \sum_{y\in G} M[x,y]\chi(y) = |G| \mathop{\mathbb{E}}_{y\in G} f(xy^{-1})\chi(y) \\ &= |G| \mathop{\mathbb{E}}_{z\in G} f(z)\chi(z^{-1}x) = |G| \mathop{\mathbb{E}}_{z\in G} f(z)\overline{\chi(z)}\chi(x) \\ &= |G|\chi(x) \mathop{\mathbb{E}}_{z\in G} f(z)\overline{\chi(z)} = |G|\chi(x)\langle \chi, f\rangle = |G|\hat{f}_\chi\chi(x).\end{aligned}$$

□

where in the second line we did the variable change $z = xy^{-1}$.

If G is an Abeliean group the characters form an orthonormal eigenvector basis of M. This shows, in particular, that all matrices of this form (regardless what f is) commute (because they share an eigenvector basis).

Let G be a group and $S \subseteq G$. We define a graph $H(V,E)$ where $V = G$ and $(x,y) \in E$ iff $x^{-1}y \in S$. If S is closed under inverse, we get an undirected graph of degree $|S|$. Many familiar graphs are obtained this way, e.g., the cycle (over $\mathbb{Z}_n$), the cube (over $\mathbb{F}_2^n$) the mesh (with closed rows and columns, over $\mathbb{F}_n^2$) and more.

Now suppose G is Abeliean and S is closed under inverse. The adjacency matrix M of $Cay(G,S)$ is captured by 1_S, i.e., $M[x,y] = 1$ iff $x^{-1}y \in S$ iff $y^{-1}x \in S$ i.e., $M[x,y] = 1_S(xy^{-1})$. By the above lemma, the characters form an orthonormal basis for $Cay(G,S)$. The first eigenvalue is the degree $|S|$ and is obtained by the identity character $\chi_\emptyset$ that is identically 1. We see that the second largest eigenvalue in absolute value of G, $\bar{\lambda}(G)$, is $max_{S\neq\emptyset}|\hat{f}_S|$. We therefore can conclude:

Lemma 21. *Suppose G is Abeliean and $S \subseteq G$ closed under inverse. Then,*

$$\bar{\lambda}(Cay(G,S)) = Bias(S).$$

References

[1] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[2] Mihir Bellare, Don Coppersmith, JOHAN Hastad, Marcos Kiwi, and Madhu Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

[3] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of computer and system sciences*, 47(3):549–595, 1993.

[4] Oded Goldreich. Three XOR-lemmasan exposition. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 248–272. Springer, 2011.

[5] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[6] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.