Fourier analysis allows us to represent general boolean functions $f\colon \{0,1\}^n \to \mathbb{R}$ as a linear combination (over the reals) of the characters χ_a, i.e. the linear functions over $\{0,1\}$. Since the characters form an orthogonal basis in the linear space of all functions from $\{0,1\}^n$ to $\mathbb{R}$, this representation always exists and is unique.

What if we want to represent f not only in terms of linear functions over $\mathbb{F}_2$, but also allow for higher degree polynomials? Then the representation is no longer unique so we cannot describe such representations in terms of formulas for coefficients like the Fourier coefficients. Instead let's look at some interesting applications of Fourier analysis and see if they can be generalized to the higher-degree setting in a meaningful and interesting way.

Let's start with the linearity test. Recall that to test if a boolean valued function f is linear, the test chooses two random points x and y and accepts if $f(x) + f(y) = f(x + y)$. If f is linear, the test always accepts, and using Fourier analysis we showed that if f is δ-*far* from linear (i.e. f differs from all linear functions on at least a δ-fraction of inputs), then the test rejects with probability at least δ.

Can we design a similar test that accepts all polynomials of degree d, but rejects all functions which are far from degree d polynomials with noticeable probability? For inspiration, let's go back to the linearity test. The starting point of the linearity test was the equivalence of two definitions of linearity: A structural definition, which says f is linear if it has the form $f(x) = a_1x_1 + \cdots + a_nx_n$, and a behavioral definition, which says that f is linear if $f(x) + f(y) = f(x + y)$ for all pairs of inputs x and y.

A structural definition for "f is a degree d polynomial" is straightforward: f is a degree d polynomial if it can be written as a linear combination of monomials $\prod_{i\in S} x_i$, where S has size at most d. How about a behavioral definition?

1 Directional derivatives and Gowers uniformity

In calculus we learn that taking derivatives reduces the degree of polynomials. There is a similar phenomenon in $\mathbb{F}_2$, if instead of derivatives we work with their discrete analogues.

Definition 1. For $f\colon \{0,1\}^n \to \{0,1\}$ and $a \in \{0,1\}^n$, the derivative of f in direction a is the function

$$D_a f(x) = f(x + a) + f(x).$$

If f is constant, then $D_a f(x) = 0$. If f is affine, then $D_a f(x) = f(a)$, which is constant. More generally, taking a discrete derivative reduces the degree by one. So if f is a degree $(d - 1)$ polynomial, then $D_{a_1} \dots D_{a_d} f = 0$ for any choice of d directions $a_1, \dots, a_d$.

The converse is also true: If f is a $\{0,1\}$-valued function such that $D_{a_1} \dots D_{a_d} f = 0$ for all $a_1, \dots, a_d \in \mathbb{F}^n$, then f must have degree $d - 1$.

By analogy with the linearity test, we could hope to test if a function is a degree-$(d-1)$ polynomial by choosing $x, a_1, \dots, a_d$ at random and accepting if $D_{a_1} \dots D_{a_d} f(x) = 0$. The bias of this test is

called the *d-uniformity* of f:

$$\mathrm{U}_d[f] = \mathrm{E}_{x,a_1,\dots,a_d\sim\{0,1\}^n}[(-1)^{D_{a_1}\dots D_{a_d}f(x)}] = \mathrm{Ee}_{x,a_1,\dots,a_d\sim\{0,1\}^n}[D_{a_1,\dots,a_d}f(x)],$$

where $\mathrm{Ee}[\cdot]$ is shorthand notation for $\mathrm{E}[(-1)^{\cdot}]$ and $D_{a_1,\dots,a_d}$ is shorthand for $D_{a_1}\dots D_{a_d}$. So f is a degree-$(d-1)$ polynomial if and only if $\mathrm{U}_d[f] = 1$. In analogy with the linearity test, we may expect that when the d-uniformity of f is bounded away from 1, then f is far from a degree-$(d-1)$ polynomials.

To get some intuition let's look at the case $d = 2$. We can write the 2-uniformity of f as

$$\mathrm{U}_2[f] = \mathrm{Ee}[D_{u,v}f(x)] = \mathrm{Ee}[f(x) + f(x+u) + f(x+v) + f(x+u+v)].$$

This expression resembles the bias of the linearity test, so we may hope to gain some insight about its value by doing a Fourier expansion. Writing $F(x) = (-1)^{f(x)}$, we have

$$\begin{aligned}
&\mathrm{E}[F(x)F(x+u)F(x+v)F(x+u+v)] \\
&\qquad = \mathrm{E}\Big[\sum_a \hat{F}_a\chi_a(x)\sum_b \hat{F}_b\chi_b(x+u)\sum_c \hat{F}_c\chi_c(x+v)\sum_d \hat{F}_d\chi_d(x+u+v)\Big] \\
&\qquad = \sum_{a,b,c,d} \hat{F}_a\hat{F}_b\hat{F}_c\hat{F}_d\,\mathrm{E}[\chi_a(x)\chi_b(x+u)\chi_c(x+v)\chi_d(x+u+v)] \\
&\qquad = \sum_{a\in\{0,1\}^n} \hat{F}_a^4
\end{aligned}$$

because the only nonvanishing expectations are those where $a = b = c = d$. By Parseval's identity, the last expression is at most $\hat{F}_a^2$, so if $\mathrm{U}_d[f] \geq 1-\delta$, then F must be $\sqrt{1-\delta}$-close to χ_a or $-\chi_a$. In other words, f is $\sqrt{1-\delta}$-close to some affine function.

Can we extend this argument to larger values of d? It is tempting to try and work out a formula with Fourier coefficients, but even for $d = 3$ it is not easy to make sense of this formula and understand how it relates to degree 2 polynomials. Instead we will do a combinatorial analysis.

2 Analysis of the low-degree test

This analysis of the Gowers low-degree test is by Bhattarchaya, Kopparty, Schoenbeck, Sudan, and Zuckerman. Their analysis concerns a slight variation of the Gowers test: They insist that the directions $a_1,\dots,a_d$ are *linearly independent*, but otherwise random. For large n, the probability of a linear dependency between goes to zero, so the rejection probability of the two tests is asymptotically the same.

Degree-$(d-1)$ test T_n: Given a function $f\colon\{0,1\}^n \to \{0,1\}$, choose $x \sim \{0,1\}^n$ at random and $a_1,\dots,a_d$ at random from $\{0,1\}^n$ provided they are linearly independent. If $D_{a_1,\dots,a_d}f(x) = 0$ accept, otherwise reject.

We can write this test more explicitly by expanding the formula for directional derivatives:

$$D_{a_1,\dots,a_d}f(x) = \sum_{c\in\{0,1\}^d} f(x+ca) \mod 2$$

where $x + ca = x + c_1a_1 + \cdots + c_da_d$.

Here we assume that $n \geq d$; otherwise, any function can be represented as a degree-n polynomial so the test can always accept. Clearly if f is a degree-$(d-1)$ polynomial, T_d always accepts.

Theorem 2. *There exist constants α_0, α_1 such that for every $\delta > 0$, if f is δ-far from all degree-$(d-1)$ polynomials, then T_n rejects f with probability at least $\min\{\alpha_0, \alpha_1\delta 2^d\}$.*

We do the analysis by induction on the number of inputs n. We'll worry about the base case later. Let's make the inductive hypothesis that if f is a function in $n-1$ inputs that is at least δ_{n-1}-far from all degree-$(d-1)$ polynomials, T_{n-1} rejects f with probability at least ρ_{n-1} and see how the analogous quantities ρ_n and δ_n for T_n relate to ρ_{n-1} and δ_{n-1}.

We fix a degree-d polynomial f in n inputs that is δ_n-far from random. We want to lower bound the probability ρ_n that T_n rejects f. When $d > n$, we can view the choice of $a_1, \ldots, a_d$ as happening in two stages: First, we choose a random $(n-1)$-dimensional affine subspace S of $\{0,1\}^n$, then we choose $a_1, \ldots, a_d$ uniformly at random from S. So the rejection probability of $T_n(f)$ can be written as the average of the rejection probabilities $T_{n-1}(f|_S)$:

$$\Pr[T_n \text{ rejects } f] = \mathrm{E}_S \Pr[T_{n-1} \text{ rejects } f|_S]$$

where $f|_S$ is the restriction of f on S, which we can identify with $\{0,1\}^{n-1}$ under an appropriate change of basis (the test is independent of the choice of basis).

We will argue as follows. If $f|_S$ is δ_{n-1}-close to a degree-$(d-1)$ polynomial on at least K of the affine subspaces S, we will argue that f itself must be δ_n-close to a degree-$(d-1)$ polynomial. (The appropriate choice of K will come out of the calculation.) Since it is not, we get that

$$\rho_n \geq \Big(1 - \frac{K}{2^n}\Big)\rho_{n-1}$$

as there are $2^{n+1} - 2 \geq 2^n$ possible choices for the subspace S.

So now let's assume that f_S is δ_{n-1}-close to a degree-$(d-1)$ polynomial on at least K of the affine subspaces S – let's call them $S_1, \ldots, S_K$ – and see what we can say about f. We will need a variant of the Schwarz-Zippel lemma for polynomials over $\mathbb{F}_2$:

Lemma 3. *If $p(x_1, \ldots, x_n)$ is a nonzero degree d polynomial over $\mathbb{F}_2$, then $\Pr_x[p(x) \neq 0] \geq 2^{-d}$.*

Now let's look at any two subspaces S_i, S_j. We know $f|_{S_i}$ and $f|_{S_j}$ are each δ_{n-1}-close to degree-$(d-1)$ polynomials p_i and p_j defined over S_i and S_j respectively. We claim that $p_i(x) = p_j(x)$ for every x in $S_i \cap S_j$: Otherwise, $p_i - p_j$ would be a nonzero degree-$(d-1)$ polynomial which is $4\delta_{n-1}$-close to $f|_{S_i} - f|_{S_j} = 0$ on $S_i \cap S_j$, which is impossible as long as $\delta_{n-1} \leq 2^{-(d+1)}$.

This says the degree-$(d-1)$ polynomials that are close to f in various subspaces are all consistent with one another. The following lemma (which you can try to prove) says that we can "glue" all this information together into a single degree-$(d-1)$ polynomial:

Lemma 4. *Let $S_1, \ldots, S_K$ be a collection of $(d-1)$-dimensional subspaces of $\{0,1\}^n$ and p_i be a degree-$(d-1)$ polynomial on S_i such that p_i and p_j agree on $S_i \cap S_j$ for every pair (i,j). If $K > 2^{d-1}$ there exists a single degree-$(d-1)$ polynomial p on $\{0,1\}^n$ that agrees with p_i on S_i for every i.*

So if $K > 2^{d-1}$ and $\delta_{n-1} \leq 2^{-(d+1)}$, we get a single polynomial p of degree $d-1$ that is δ_{n-1}-close to f on all of $S_1, \ldots, S_K$. We show that p has to be $2\delta_{n-1} + 4/K$-close to f.

To show this, we can assume none of the subspaces S_i, S_j are complementary, for otherwise the claim is trivial. Otherwise let $N_i(x)$ be an indicator for the event $x \in S_i$, $1 \leq i \leq K$. The random variables $N_i(x)$ when x is chosen at random have mean $K/2$ and are pairwise independent, so by Chebyshev's inequality

$$\Pr[N_1(x) + \cdots + N_K(x) < K/4] \leq 4/K.$$

Say x is good if $N_1(x) + \cdots + N_K(x) \geq K/4$. Then the probability that x is good *and* $f(x) \neq p(x)$ is at least $\delta_n - 4/K$. For a random choice of i, we get that

$$\Pr_{i,x}[x \in S_i \text{ and } f(x) \neq p(x)] \geq (\delta_n - 4/K)/4$$

so for at least one i, the distance between $f(x)$ and $p(x)$ on S_i must be at least $(\delta_n - 4/K)/2$, and so δ_{n-1} must be at least this large.

Setting $K = 1/\delta_{n-1}$, we conclude that if $\delta_{n-1} \leq 2^{-(d+1)}$ and T_{n-1} rejects functions that are at least δ_{n-1}-far from having degree $d-1$ with probability at least ρ_{n-1}, then T_n rejects functions that are at least $6\delta_{n-1}$-far from having degree $d-1$ with probability at least $\rho_n = (1 - 1/\delta_{n-1}2^n)\rho_{n-1}$.

We now sketch the proof of the theorem. The inductive statement we will prove is that T_n rejects all functions that are at least δ-far from degree $(d-1)$ with probability at least $\min\{\varepsilon_n/24, \varepsilon_n\delta 2^d\}$ where $1/2 \geq \varepsilon_n \geq \varepsilon_{n-1}(1 - 2^{-n+d+1})$.

Suppose f is δ-far from having degree $(d-1)$. First suppose $\delta \leq 2^{-d}/2$ and let p be the closest degree $d-1$ polynomial to f. Using the expanded form of the directional derivative, we have

$$\begin{aligned}
&\Pr[D_{a_1,\ldots,a_d}p(x) \neq 0] \\
&\quad \geq \Pr[f(x+ca) = p(x+ca) \text{ for exactly one } c] \\
&\quad = 2^d \Pr[f(x) = p(x) \text{ and } f(x+ca) \neq p(x+ca) \text{ for all } c \neq 0] \\
&\quad = 2^d \Pr[f(x) = p(x)] \Pr[f(x) = p(x) \text{ and } f(x+ca) \neq p(x+ca) \text{ for all } c \neq 0 \mid f(x) = p(x)] \\
&\quad \geq \Pr[f(x) = p(x)]\Big(1 - \sum\nolimits_{c \neq 0} \Pr[f(x+ca) \neq p(x+ca) \mid f(x) = p(x)]\Big) \\
&\quad = 2^d\delta(1 - (2^d - 1)\delta_n \cdot 2^n/(2^n - 1)) \geq 2^{d-1}\delta.
\end{aligned}$$

In the second to last line we used the fact that for every nonzero c, the points x and $x + ca$ are two random distinct points in $\{0,1\}^n$. In this case the rejection probability is at least $2^d\delta/2$ and we are done.

Now suppose $\delta > 2^{-d}/2$. By the above analysis, in this case the test rejects with probability at least $(1 - 2^{-n+d+1})\rho_{n-1}$, where ρ_{n-1} is the rejection probability for T_{n-1} at distance $\min\{\delta/6, 2^{-(d+1)}\} > 2^{-d}/12$. By inductive hypothesis, $\rho_{n-1} \geq \min\{\varepsilon_{n-1}/24, \varepsilon_{n-1}(\delta/6)2^{d-1}\} = \varepsilon_{n-1}/24$, and we conclude that $\varepsilon_n \geq \varepsilon_{n-1}(1 - 2^{-n+d+1})$ as desired.

Now we can iterate this statement to obtain

$$\varepsilon_n \geq \varepsilon_{d+3}(1 - 1/4)(1 - 1/8)\cdots \geq \varepsilon_{d+3}/2$$

With a little bit more work it is possible to show that $\varepsilon_{d+3} \geq 1/8$, providing a base case and proving the theorem.

3 Low correlation with all low-degree polynomials

Let $\mathcal{F}$ be a family of functions from $\{0,1\}^n$ to $\{1,-1\}$ and $G\colon \{0,1\}^n \to \{1,-1\}$ be a function outside $\mathcal{F}$. The *correlation* between $\mathcal{F}$ and G is the maximum value of $\mathrm{E}[F(x)G(x)]$ over all F in $\mathcal{F}$.

For example, if $\mathcal{F}$ is the family of circuits of size s, then if we find a G that has correlation, say, at most 0.9 with $\mathcal{F}$ we have proved that no circuit of size s can compute G on more than a 0.8 fraction of inputs. Finding such explicit G is a difficult task, but using the tools from today's lecture we can construct functions that have very low correlation with all polynomials of degree $d-1$ when d is small compared to n.

Let's start with linear functions. We are looking for a function g that has small correlation with all linear functions or characters χ_a and their complements. Since $\mathrm{E}[G(x)\chi_a(x)] = \hat{G}_a$, we are looking for a function whose maximum Fourier coefficient is as small as possible in absolute value.

By Parseval's identity we know that $\sum_a \hat{G}_a^2 = 1$, so there must exist a Fourier coefficient of absolute value at least $2^{-n/2}$. Moreover, if this value is attained, then all Fourier coefficient must have the same absolute value. Can we construct a g such that $|\hat{G}_a| = 2^{-n/2}$ for all a? When n is even, one example is the inner product function $G(x) = (-1)^{g(x)}$

$$g(x) = x_1x_2 + x_3x_4 + \cdots + x_{n-1}x_n.$$

One way to see that $|\hat{G}_a| = 2^{-n/2}$ for all a is like this. A calculation shows that the Fourier coefficients of $(-1)^{xy}$ all have absolute value $1/2$. The Fourier coefficients of a product of functions over disjoint inputs are the product of the Fourier coefficients of the respective functions. Since $G(x)$ is the product of $n/2$ copies of $(-1)^{xy}$ over disjoint inputs, all its Fourier coefficients must have absolute value $2^{-n/2}$.

For larger d, we do not know in general what are the functions that minimize the correlation with degree $(d-1)$ polynomials over $\mathbb{F}_2$, but we can give examples of functions with correlation as small as $2^{-\Omega(n)}$ for any fixed d. One example is provided by the natural generalization of the inner product to larger degrees:

Theorem 5. *When n is a multiple of d, the function $G(x) = (-1)^{g(x)}$, where*

$$g(x) = x_1x_2\ldots x_d + \cdots + x_{n-d+1}x_{n-d+2}\ldots x_n$$

has correlation at most $2^{-\Omega(n/d2^d)}$ with the family of degree $(d-1)$ polynomials over $\mathbb{F}_2$.

To prove this theorem, we will use the d-uniformity of g to bound the correlation between G and the family of degree-$(d-1)$ polynomials. This is due to two very nice properties of U_d. Let g be the $\{0,1\}$-valued counterpart of G, so that $G(x) = (-1)^{g(x)}$. The correlation between G and a polynomial p of degree at most $d-1$ is given by $\mathrm{E}[G(x)(-1)^{p(x)}] = \mathrm{Ee}[g(x)+p(x)]$; this is the quantity we want to bound. Instead of expectations, we will work with d-uniformity. The first observation is that

$$\mathrm{U}_d[g+p] = U_d[g] \text{ for every polynomial } p \text{ of degree at most } d-1$$

because derivatives are linear and taking d derivatives of a degree $d-1$ polynomial make the

polynomial vanish. The second one is the sometimes called *Cauchy-Schwarz-Gowers inequality*:

$$\begin{aligned}
\mathrm{U}_{d-1}[g]^2 &= \mathrm{Ee}[D_{a_1,\ldots,a_{d-1}} g(x)]^2 \\
&= \mathrm{E}_{a_1,\ldots,a_{d-1}} \left[\mathrm{Ee}_x[D_{a_1,\ldots,a_{d-1}} g(x)]\right]^2 \\
&\leq \mathrm{E}_{a_1,\ldots,a_{d-1}} \left[\mathrm{Ee}_x[D_{a_1,\ldots,a_{d-1}} g(x)]^2\right] \\
&= \mathrm{E}_{a_1,\ldots,a_{d-1}} \left[\mathrm{Ee}_{x,y}[D_{a_1,\ldots,a_{d-1}} g(x) + D_{a_1,\ldots,a_{d-1}} g(y)]\right] \\
&= \mathrm{Ee}_{a_1,\ldots,a_{d-1},x,y}[D_{a_1,\ldots,a_{d-1},x+y} g(x)] \\
&= \mathrm{U}_d[g]
\end{aligned}$$

Iterating this inequality we get that

$$\mathrm{Ee}[g]^{2^d} = \mathrm{U}_1[g]^{2^d} \leq \mathrm{U}_d[g].$$

Putting the two observations together, we get that the correlation of g with degree $d-1$ polynomials is upper bounded by $\mathrm{U}_d[g]^{2^{-d}}$.

To prove theorem 5 all we need to do is calculate $U_d[g]$. Just like Fourier coefficients, uniformity measures are multiplicative for functions over disjoint inputs, so we have

$$\mathrm{U}_d[g] = \mathrm{U}_d[x_1 x_2 \ldots x_d]^{n/d}.$$

I don't know how to calculate $\mathrm{U}_d[x_1 x_2 \ldots x_d]$, but by the Schwartz-Zippel lemma we know that

$$\Pr[p(x) \neq x_1 x_2 \ldots x_d] \geq 2^{-d}$$

for every polynomial p of degree $d-1$ or less. So the function $x_1 \ldots x_d$ is 2^{-d}-far from all degree-d polynomials. By Theorem 2, $D_{a_1 \ldots a_d} x_1 \ldots x_d$ is nonzero with at least constant probability, so

$$\mathrm{U}_d[x_1 x_2 \ldots x_d] \leq 1 - \alpha \quad \text{for some } \alpha > 0.$$

It follows that the correlation between G and degree $d-1$ polynomials is at most $(1-\alpha)^{n/d2^d}$ as promised.