

Recall that an  $(N, K)$  two-source hitter is a function  $f: [N]^2 \rightarrow \{0, 1\}$  that is not constant on any product of sets  $S \times T$  of size  $K$  each. We showed  $(N, 2 \log N + 1)$  two-source hitters exist, but even after optimizations the best construction of such a hitter took us around  $2^{n^2}$  time, where  $N = 2^n$ .

Here is a simple proposal for a two-source hitter, due to Chor and Goldreich:

$$f(x, y) = \langle x, y \rangle \bmod 2 = x_1y_1 + x_2y_2 + \cdots + x_ny_n \bmod 2$$

where  $x_i$  and  $y_i$  are the  $i$ th bit of  $x$  and  $y$ , respectively.

How well does this work? It can never work too well:  $S$  and  $T$  must have size at least  $\sqrt{N}$  (at least when  $n$  is even; can you see why)? However, if  $S$  and  $T$  are larger than that, we have

**Theorem 1.** *The function  $\langle x, y \rangle \bmod 2$  is a  $(N, \sqrt{N} + 1)$  two-source hitter.*

Proving this theorem from scratch is not impossible, although I suspect it would be a bit challenging. Using Fourier analysis, the answer comes out staring at us.

## 1 Fourier analysis

Fourier analysis allows us to (literally) look at Boolean functions from a different angle. Instead of Boolean functions, we look at more general functions that take inputs from  $\{0, 1\}^n$ , but produce a real-valued output. A Boolean function is the special case when there are only two possible output values among all reals, usually the values  $\{0, 1\}$  or  $\{-1, 1\}$ .

We can think of a function  $f$  from  $\{0, 1\}^n$  to the real numbers as a vector in  $2^n$  dimensional space. The standard basis for this space consists of those vectors that have entry 1 in exactly one of the  $2^n$  positions, and entry 0 everywhere else. These are the  $2^n$  functions  $\delta_a$  ( $a \in \{0, 1\}^n$ ), which take value 0 everywhere except at  $a$ , where they take value 1.

Now any function  $f$  over  $\{0, 1\}^n$  can be written as a linear combination of the functions  $\delta_a$ . The coefficient in front of  $\delta_a$  – the value  $f(a)$  – is then simply just the inner product between  $f$  and  $\delta_a$ .

The Fourier transform allows us to express  $f$  in a different orthogonal basis. This is the basis consisting of the character functions  $\chi_a: \{0, 1\}^n \rightarrow \mathbb{R}$

$$\chi_a(x) = (-1)^{\langle x, a \rangle}, \quad a \in \{0, 1\}^n.$$

It is easy to verify that, viewed as vectors, the functions  $\chi_a$  are orthogonal and each has norm  $2^{n/2}$ . Instead of normalizing these functions, it is more convenient to work with the normalized inner product

$$\text{inner product of } f \text{ and } g = \mathbb{E}_{x \sim \{0, 1\}^n} [f(x)g(x)] = \frac{1}{2^n} \sum_{x \in \{0, 1\}^n} f(x)g(x).$$

In this basis, a function  $f: \{0, 1\}^n \rightarrow \mathbb{R}$  can be written as a linear combination of the  $\chi_a$ s:

$$f(x) = \sum_{a \in \{0, 1\}^n} \hat{f}_a \cdot \chi_a(x)$$

where the coefficient  $\hat{f}_a$  is just the inner product of  $f$  and  $\chi_a$ :

$$\hat{f}_a = \mathbb{E}_{x \sim \{0,1\}^n} [f(x)\chi_a(x)]. \quad (1)$$

**Parseval's identity** Since the bases  $\{\delta_a\}$  and  $\{\chi_a\}$  are related by an orthonormal transformation (up to the normalization factor  $2^{n/2}$ ), the sum of the squares of the coefficients of  $f$  in these two bases should be the same. This is Parseval's identity:

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

We can also prove this algebraically by a calculation:

$$\begin{aligned} \mathbb{E}_{x \sim \{0,1\}^n} [f(x)^2] &= \mathbb{E}_{x \sim \{0,1\}^n} \left[ \left( \sum_{a \in \{0,1\}^n} \hat{f}_a \chi_a(x) \right)^2 \right] \\ &= \mathbb{E}_{x \sim \{0,1\}^n} \left[ \sum_{a,b \in \{0,1\}^n} \hat{f}_a \hat{f}_b \chi_a(x) \chi_b(x) \right] \\ &= \sum_{a,b \in \{0,1\}^n} \hat{f}_a \hat{f}_b \mathbb{E}_{x \sim \{0,1\}^n} [\chi_a(x) \chi_b(x)] \\ &= \sum_{a \in \{0,1\}^n} \hat{f}_a^2 \end{aligned}$$

as the only surviving terms in the summation over  $a$  and  $b$  are those where  $a = b$ .

If  $f$  takes values in the range  $[-1, 1]$ , we obtain that the sum of the squares of the Fourier coefficients is at most 1. If  $f$  takes only the values  $-1$  and  $1$ , then the sum of the squares is exactly 1, and so the squares of the Fourier coefficients can be thought of as a probability distribution over  $\{0, 1\}^n$ .

## 2 Analysis of the Chor-Goldreich construction

Let us see immediately how this can be used to analyze the Chor-Goldreich construction. The first step in doing Fourier analysis is to come up with an algebraic representation of the problem we are looking at. In the Chor-Goldreich construction, we want to say that there is some variation among the values  $\langle x, y \rangle$ , where  $x$  and  $y$  range from  $S$  and  $T$  respectively. One way to say this is that when  $x$  and  $y$  are chosen at random from their respective sets, the expected value  $\mathbb{E}[\langle x, y \rangle]$  is not identically zero or identically one:

$$0 < \mathbb{E}_{x \sim S, y \sim T} [\langle x, y \rangle \bmod 2] < 1.$$

Instead of working with a separate lower bound and upper bound, it is more convenient to just have one. We can accomplish this by the usual "change of constants" from  $\{0, 1\}$  to  $\{1, -1\}$ . In this notation we want to show that:

$$|bias| < 1 \quad \text{where} \quad bias = \mathbb{E}_{x \sim S, y \sim T} [(-1)^{\langle x, y \rangle}].$$

Now comes a useful trick in Fourier analysis: We convert sets into functions. In our case, we will replace  $S$  and  $T$  by their respective *indicator functions*:

$$f(x) = \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{otherwise} \end{cases} \quad g(y) = \begin{cases} 1, & \text{if } y \in T \\ 0, & \text{otherwise} \end{cases}$$

We can write:

$$\begin{aligned}
bias &= \frac{1}{|S| \cdot |T|} \sum_{x,y \in \{0,1\}^n} f(x)g(y)(-1)^{\langle x,y \rangle} \\
&= \frac{2^n}{|S| \cdot |T|} \sum_{x \in \{0,1\}^n} f(x) \mathbb{E}_{y \sim \{0,1\}^n} [g(y)(-1)^{\langle x,y \rangle}] \\
&= \frac{2^n}{|S| \cdot |T|} \sum_{x \in \{0,1\}^n} f(x) \cdot \hat{g}_x.
\end{aligned}$$

Another salient feature of these proofs is the Cauchy-Schwarz inequality.

$$\sum_{x \in \{0,1\}^n} f(x) \cdot \hat{g}_x \leq \sqrt{\sum_{x \in \{0,1\}^n} f(x)^2} \cdot \sqrt{\sum_{x \in \{0,1\}^n} \hat{g}_x^2}.$$

By definition, the first term equals  $\sqrt{|S|}$ . By Parseval's identity, the second term equals  $\sqrt{|T|/2^n}$  and

$$bias \leq \sqrt{2^n/|S||T|}$$

so if  $|S|, |T| > 2^{n/2}$ , as we assumed, then  $\langle x, y \rangle$  is indeed a two-source hitter.

### 3 The linearity test

There are two different ways of saying that a function  $f: \{0, 1\}^n \rightarrow \{0, 1\}$  is linear:

1.  $f$  is linear if it has the form  $f(x) = a_1x_1 + \dots + a_nx_n$ .
2.  $f$  is linear if  $f(x + y) = f(x) + f(y)$  for all  $x, y \in \{0, 1\}^n$ .

It is clear that the first characterization implies the second. What about the other way? You can argue this direction with some algebra. But there is an elegant proof via Fourier analysis that has nice extensions relevant in computer science.

So let's assume that  $f(x + y) = f(x) + f(y)$  for all  $x, y \in \{0, 1\}^n$  and see what we can deduce about  $f$  itself. To apply Fourier analysis, we replace the 0, 1 outputs of  $f$  by 1, -1 via the usual substitution  $F(x) = (-1)^{f(x)}$ . Then the assumption becomes  $F(x + y) = F(x)F(y)$ , or

$$F(x)F(y)F(x + y) = 1 \quad \text{for all } x, y \in \{0, 1\}^n.$$

Averaging over all pairs  $x, y$ , we get

$$\mathbb{E}_{x,y \sim \{0,1\}^n} [F(x)F(y)F(x + y)] = 1. \tag{2}$$

Fourier analysis is often helpful with expressions that look like the one on the left. Let's see what we get:

$$\begin{aligned}
\mathbb{E}[F(x)F(y)F(x + y)] &= \mathbb{E}_{x,y \sim \{0,1\}^n} \left( \sum_{a \in \{0,1\}^n} \hat{F}_a \chi_a(x) \right) \left( \sum_{b \in \{0,1\}^n} \hat{F}_b \chi_b(y) \right) \left( \sum_{c \in \{0,1\}^n} \hat{F}_c \chi_c(x + y) \right) \\
&= \mathbb{E}_{x,y \sim \{0,1\}^n} \sum_{a,b,c \in \{0,1\}^n} \hat{F}_a \hat{F}_b \hat{F}_c \chi_a(x) \chi_b(y) \chi_c(x + y) \\
&= \sum_{a,b,c \in \{0,1\}^n} \hat{F}_a \hat{F}_b \hat{F}_c \cdot \mathbb{E}_{x,y \sim \{0,1\}^n} [\chi_a(x) \chi_b(y) \chi_c(x + y)].
\end{aligned}$$

Recalling that  $\chi_a(x) = (-1)^{\langle a, x \rangle}$ , we can write

$$\mathbb{E}_{x, y \sim \{0,1\}^n} [\chi_a(x) \chi_b(y) \chi_c(x+y)] = \mathbb{E}_{x, y \sim \{0,1\}^n} [\chi_{a+c}(x) \chi_{b+c}(y)] = \mathbb{E}_{x \sim \{0,1\}^n} [\chi_{a+c}(x)] \mathbb{E}_{y \sim \{0,1\}^n} [\chi_{b+c}(y)].$$

Now notice that this expression equals zero unless  $a = c$  and  $b = c$ , and so

$$\mathbb{E}[F(x)F(y)F(x+y)] = \sum_{\substack{a, b, c \in \{0,1\}^n: \\ a = c \text{ and } b = c}} \hat{F}_a \hat{F}_b \hat{F}_c = \sum_{a \in \{0,1\}^n} \hat{F}_a^3.$$

Therefore, from (2) we must have  $\sum \hat{F}_a^3 = 1$ . But by Parseval's identity  $\sum \hat{F}_a^2 = 1$ ; this is only possible if  $\hat{F}_a = 1$  for exactly one value of  $a$ , which means that  $F(x) = \chi_a(x)$ .

One important advantage of this Fourier-analytic proof is that it is *robust*: Even if the constraint  $f(x) + f(y) = f(x+y)$  fails to hold sometimes, we cannot say anymore that  $f$  is linear, but we can still say that it is “close” to linear.

**Theorem 2.** *Suppose that  $f(x) + f(y) = f(x+y)$  with probability  $1 - \delta$  when  $x$  and  $y$  are chosen independently at random from  $\{0, 1\}^n$ . Then  $\Pr[f(x) = \langle a, x \rangle \bmod 2] \geq 1 - \delta$  for some  $a \in \{0, 1\}^n$ .*

*Proof.* Writing  $F(x) = (-1)^{f(x)}$ , we have

$$\mathbb{E}_{x, y \sim \{0,1\}^n} [F(x)F(y)F(x+y)] = \Pr[F(x)F(y)F(x+y) = 1] - \Pr[F(x)F(y)F(x+y) = -1] = 1 - 2\delta$$

and so

$$\sum_{a \in \{0,1\}^n} \hat{F}_a^3 = 1 - 2\delta.$$

Since  $\sum \hat{F}_a^3 \leq (\max \hat{F}_a) \sum \hat{F}_a^2 = \max \hat{F}_a$ , there must exist  $a \in \mathbb{F}^n$  such that  $\hat{F}_a \geq 1 - 2\delta$ . But  $\hat{F}_a = \mathbb{E}[F(x)\chi_a(x)] = \Pr[F(x) = \chi_a(x)] - \Pr[F(x) \neq \chi_a(x)]$ , so  $\Pr[F(x) = \chi_a(x)] \geq 1 - \delta$ .  $\square$

This theorem can be interpreted as a statement about the following scenario. We are given an unknown function  $f$  and we want to test if  $f$  is a linear function. A natural idea is to choose two random inputs  $x$  and  $y$  and check if  $f(x) + f(y) = f(x+y)$ . If  $f$  is linear, the test will certainly pass. The theorem tells us that if  $f$  passes the test with high probability, then we can be sure that  $f$  is close to a linear function.