The Sarnak Conjecture: Orthogonality of the Mobius Function on Bounded Depth Circuits

Chapter 1

The Sarnak Conjecture. A Basic Case.

The Mobius function is defined, as usual:

$$\mu(n) = \begin{cases} 1 & n = 1 \\ (-1)^k & n = p_1 p_2 \dots p_k \\ 0 & \text{otherwise} \end{cases}$$

The analytic behavior of this function is very important, with regard to prime numbers. The specific question of "orthogonality" with an integer sequence $F(n), n \geq 1$ is important, since many questions in analytic number theory are equivalent to particular cases of such orthogonality.

More specifically, we define the "orthogonality" of μ and F as the analytic behavior:

$$\sum_{n \leq x} \mu(n) F(n) = o(x) \text{ as } x \to \infty$$

A function F which is orthogonal on μ is called *deterministic*, and *nondeterministic* otherwise.

The main conjecture regarding the question of orthogonality is stated by Sarnak, that this is true for functions that are well-behaved, in some sense which shall be defined in the next section. It has been proven for some specific classes of functions, while the question is still open for some other classes.

A comprehensive list of classes for which the conjecture has been proven (in some cases very recently) can be found in [2].

- If $F(n) \equiv 1$, then the relation becomes:

  $$\sum_{n \leq x} \mu(n) = o(x)$$

  which is equivalent to the prime number theorem. We shall prove this equivalence in the last section of this chapter.

- If $F(n)$ is a periodic function, then the relation is equivalent to the prime number theorem in arithmetic sequences, or Dirichlet's theorem.

- For $F(n)$ a function computable by bounded depth circuits AC^0 (to be defined later on), the fact that Sarnak's conjecture holds is a recent result by Ben Green and the main focus of the thesis.

- $F(n) = \mu(n)$ is nondeterministic; namely, we have:

  $$\frac{1}{N} \sum_{n \leq N} \mu^2(n) \to \frac{6}{\pi^2} \tag{1.1}$$

The proof for 1.1 can be found in [10], the following is an abridged version of it. The main idea is that the asymptotic limit $\frac{1}{N}\sum_{x\leq N}\mu^2(x)$ represents the probability that a number is squarefree, that is, it does not belong to any arithmetic progression of ratio p^2. Formally, let χ_{p^2} be the characteristic function for this progression. Then, use the cutoff of the Taylor expansion series:

$$\begin{aligned}\mu^2(n) &= \prod_p (1-\chi_{p^2}(n)) \\ &= 1-\sum_{p_1}\chi_{p_1^2}(n)+\sum_{p_1<p_2}\chi_{p_1^2}(n)\chi_{p_2^2}(n)-\cdots+(-1)^k\sum_{p_1<p_2<\cdots<p_k}\chi_{p_1^2}(n)\chi_{p_2^2}(n)\ldots\chi_{p_k^2}(n)+\epsilon_k(n)\end{aligned}$$

where $\epsilon_k(n)$ is the remainder term, for which $\lim_{k\to\infty}\epsilon_k(n)=0$. Average this equality for all n:

$$\frac{1}{N}\sum_{n\leq N}\mu(n) = 1-\sum_{p_1}\frac{1}{p_1^2}+\sum_{p_1<p_2}\frac{1}{(p_1p_2)^2}-\cdots+(-1)^k\sum_{p_1<p_2<\cdots<p_k}\frac{1}{(p_1p_2\ldots p_k)^2}+\frac{\sum\limits_{n\leq N}\epsilon_k(n)}{N}$$

therefore, by passing to the limit in both N and k:

$$\lim_{n\to\infty}\frac{1}{N}\sum_{x\leq N}\mu^2(x)=\prod_p\left(1-\frac{1}{p^2}\right)$$

The product can be evaluated as well, by inverting it and breaking down the geometric series.

$$\begin{aligned}\prod_p\frac{1}{1-\frac{1}{p^2}} &= \prod_p\left(1+\frac{1}{p^2}+\frac{1}{p^4}+\ldots\right)\\ &= \sum_{n\geq 1}\frac{1}{n^2}\\ &= \frac{1}{\zeta(2)}=\frac{\pi^2}{6}\end{aligned}$$

automatically implying 1.1.

1.1 Topological entropy. The formal conjecture

This section formalizes the definition of the Sarnak conjecture.

Conjecture 1.1 *(Sarnak) The relationship:*

$$\sum_{n\leq x}\mu(n)F(n)=o(x)\ \text{ as } x\to\infty$$

holds for functions F that have zero topological entropy.

We now define the concept of topological entropy.
For a function $f:\mathbb{R}\to\mathbb{C}$ and for a given $m\in\mathbb{N}$, look at the set:

$$S_m=\{(f(n+1),f(n+1),\ldots f(n+m))|n\in\mathbb{N}\}\subset\mathbb{C}^m$$

and, for any $\epsilon>0$, define $B(m,\epsilon)$ to be the number of balls of radius ϵ under the standard metric of $\mathbb{C}$ necessary to cover S_m. Then the topological entropy σ of f is defined as:

$$\sigma=\limsup_{m\to\infty}\frac{1}{m}\log\left(B(m,\epsilon)\right)$$

In other words, $\sigma \geq 0$ is the least parameter such that, for any $\epsilon > 0$ and for $m \to \infty$, S_m can be covered by $O(\exp(\sigma m + o(m)))$ balls of radius ϵ. Intuitively, if f is Boolean, the number of m-tuples considered in counting the size of S_m is $2^m = \exp(m \log 2)$; the topological entropy σ represents the lowest factor that can replace the $\log 2$, as m goes to infinity.

A few examples:

- The constant function $f \equiv 1$ clearly has $|S_m| = 1$, and therefore has zero entropy. This case will be discussed in detail below.
- For a periodic function of period P, the set S_m can have at most P different elements. In terms of the Sarnak conjecture, this case corresponds to the PNT applied in arithmetic progressions.
- Predictably, μ does not have zero entropy; it can be calculated that its entropy is actually $\frac{6}{\pi^2} \log 2$. This is to be expected, agreeing with 1.1, yet the proof is highly nontrivial. It can be found in [11].

1.2 Prime number theorem equivalence

This section will prove that the Mobius orthogonality relation:

$$\sum_{n \leq x} \mu(n) = o(x)$$

is equivalent to the prime number theorem, stated as:

$$\sum_{p \leq x} 1 = (1 + o(1)) \frac{x}{\log(x)}$$

The following proof loosely follows the sketch set forth by Terence Tao in [2], filling in the gaps and unproven claims.

We introduce the von Mangoldt function $\Lambda(n)$, defined as:

$$\Lambda(n) = \begin{cases} \log p & n = p^k \\ 0 & \text{otherwise} \end{cases}$$

Define the following sums that will be of interest for us.

$$\pi(x) = \sum_{p \leq x} 1$$
$$\nu(x) = \sum_{p \leq x} \log(p)$$
$$\Psi(x) = \sum_{n \leq x} \Lambda(n)$$

For example, it is relatively straightforward to obtain a weaker bound on $\pi(x)$. Notice that the binomial $\binom{2n}{n}$ is smaller than 4^n and also that it is a multiple of all primes between n and $2n$. Therefore:

$$n \log 4 \geq \sum_{n \leq p \leq 2n} \log p \geq \sum_{n \leq p \leq 2n} \log n$$

$$\frac{n \log 4}{\log n} \geq \sum_{n \leq p \leq 2n} 1$$

and, ultimately, by summing this up repeatedly:

$$\log 4 \sum_{k=0}^{\log_2 n-1} \left(\frac{\frac{n}{2^k}}{\log \frac{n}{2^k}} \right) \geq \pi(n)$$

We want to prove the sum on the left is actually $O(\frac{n}{\log n})$. To do this, split it in half, and analyze each section separately:

$$\begin{aligned}
\frac{1}{\log 4}\pi(n) \leq \sum_{k=0}^{\log_2 n-1} \left(\frac{\frac{n}{2^k}}{\log \frac{n}{2^k}} \right) &= \sum_{k=0}^{\frac{\log_2 n}{2}} \left(\frac{\frac{n}{2^k}}{\log n - k \log 2} \right) + \sum_{k=\frac{\log_2 n}{2}}^{\log_2 n-1} \left(\frac{\frac{n}{2^k}}{\log n - k \log 2} \right) \\
&\leq \sum_{k=0}^{\frac{\log_2 n}{2}} \left(\frac{\frac{n}{2^k}}{\log n - \frac{\log n}{2}} \right) + \sum_{k=\frac{\log_2 n}{2}}^{\log_2 n-1} \left(\frac{\frac{n}{\sqrt{n}}}{1} \right) \\
&\leq \frac{n}{\log n} \sum_{k=0}^{\frac{\log_2 n}{2}} \frac{1}{2^k} + \sum_{k=\frac{\log_2 n}{2}}^{\log_2 n-1} \sqrt{n} \\
&\leq O\left(\frac{n}{\log n} \right) + O\left(\sqrt{n} \log n \right)
\end{aligned}$$

In conclusion:

$$\pi(n) = O\left(\frac{n}{\log n} \right) \tag{1.2}$$

which is a weaker version of the prime number theorem:

$$\pi(n) = \frac{n}{\log n} + o\left(\frac{n}{\log n} \right)$$

1.2.1 Restating the problem

We shall prove that π and Ψ are essentially proportional to each other, thus allowing us to restate the prime number theorem, namely that:

$$\pi(x) = (1 + o(1))\frac{x}{\log(x)} \iff \Psi(x) = (1 + o(1))x$$

For one side of the asymptotic inequality, break down Ψ as follows:

$$\Psi(x) = \sum_{n \leq x} \Lambda(n) = \sum_{p \leq x,\ p^\alpha < x < p^{\alpha+1}} \alpha \log(p) = \sum_{p \leq x} \log p^\alpha \leq \log(x)\pi(x)$$

For the other side, write as follows:

$$\begin{aligned}
\Psi(x) \geq \nu(x) &= \sum_{k=0}^{\log_2 x} \sum_{\frac{x}{2^{k+1}} \leq p \leq \frac{x}{2^k}} \log p \\
&\geq \sum_{k=0}^{\log_2 x} \sum_{\frac{x}{2^{k+1}} \leq p \leq \frac{x}{2^k}} \log(x) - (k+1)\log 2 \\
&\geq \pi(x)\log(x) - \log 2 \sum_{k=0}^{\log_2 x} (k+1) \sum_{\frac{x}{2^{k+1}} \leq p \leq \frac{x}{2^k}} 1 \\
&\geq \pi(x)\log(x) - \log 2 \sum_{k=0}^{\log_2 x} (k+1) \left(\pi\left(\frac{x}{2^k}\right) - \pi\left(\frac{x}{2^{k+1}}\right)\right) \\
&\geq \pi(x)\log(x) - \log(2) \sum_{k=0}^{\log_2 x} \pi\left(\frac{x}{2^k}\right)
\end{aligned}$$

Pick $0 \leq A \leq \log_2 x$, and split up the sum in two pieces: up to A, and beyond A. Then, by using the simpler estimation (1.2) that we proved $\pi(x) = O(\frac{x}{\log x})$ we can deduce:

$$\begin{aligned}
\sum_{k=0}^{\log_2 x} \pi\left(\frac{x}{2^k}\right) &= \sum_{k=0}^{A} \pi\left(\frac{x}{2^k}\right) + \sum_{k=A+1}^{\log_2 x} \pi\left(\frac{x}{2^k}\right) \\
&\leq A \cdot O\left(\frac{x}{\log x}\right) + (\log_2 x - A) O\left(\frac{\frac{x}{2^A}}{\log x - A\log 2}\right)
\end{aligned}$$

By picking a suitable A, we can ensure that this function is $o(x)$. For example, pick $A = \log\log x$:

$$\begin{aligned}
\sum_{k=0}^{\log_2 x} \pi\left(\frac{x}{2^k}\right) &\leq \log\log x \cdot O\left(\frac{x}{\log x}\right) + (\log_2 x - \log\log x) O\left(\frac{\frac{x}{\log x}}{\log x - \log\log x \log 2}\right) \\
&\leq O\left(\frac{x\log\log x}{\log x}\right) + O\left(\frac{\frac{x}{\log x}\log x}{\log x}\right) \\
&\leq o(x)
\end{aligned}$$

Therefore the other inequality is proven, and we can conclude:

$$\pi(x)\log(x) \geq \Psi(x) \geq \pi(x)\log(x) - o(x)$$

and the equivalence is now straightforward, via (1.2):

$$\boxed{\pi(x) = (1+o(1))\frac{x}{\log(x)} \iff \Psi(x) = (1+o(1))x}$$

1.2.2 Assuming the PNT

Start with the identity:

$$\boxed{-\mu(n)\log(n) = \sum_{d|n} \mu(d)\Lambda(\frac{n}{d})} \tag{1.3}$$

This identity comes from a double application of the Mobius inversion formula. The formula itself states that, for f, g two arithmetic functions, we have the following property:

$$f(n) = \sum_{d|n} g(d) \iff g(n) = \sum_{d|n} \mu(d) f(n/d), \ \forall n \in \mathbb{N}$$

A proof, and further detaliation of this basic theorem in number theory can be found in many sources, such as [12]. For this specific application, the process follows:

$$\begin{aligned}
\log(n) = \sum_{d|n} \Lambda(d) &\implies \Lambda(n) = \sum_{d|n} \mu(d) \log(\frac{n}{d}) \\
&\implies \Lambda(n) = \log n \sum_{d|n} \mu(d) + \sum_{d|n} -\mu(d) \log d \\
&\implies \Lambda(n) = \sum_{d|n} -\mu(d) \log d \\
&\implies -\mu(n) \log(n) = \sum_{d|n} \mu(d) \Lambda(\frac{n}{d})
\end{aligned}$$

Sum (1.3) for $1 \leq n \leq x$:

$$-\sum_{n \leq x} \mu(n) \log(n) = \sum_{d \leq x} \mu(d) \sum_{m \leq x/d} \Lambda(m) = \sum_{d \leq x} \mu(d) \Psi(\frac{x}{d})$$

Let us evaluate the behavior of the right term. Assuming that we know the prime number theorem, that is, for any ϵ and sufficiently large x/d:

$$\Psi(\frac{x}{d}) = (1 + o(1))\frac{x}{d} = (1 + O(\epsilon))\frac{x}{d}$$

the right term evaluates to:

$$\left| \sum_{d \leq x} \mu(d) \frac{x}{d} + \mu(d) o(1) \frac{x}{d} \right| \leq x \left| \sum_{d \leq x} \frac{\mu(d)}{d} \right| + x \cdot o(1) \sum_{d \leq x} \frac{1}{d}$$

The unknown sum can be evaluated by summing the identity $1_{n=1} = \sum_{d|n} \mu(d)$ for $n \leq x$ and observing:

$$1 = \sum_{d \leq x} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d \leq x} \mu(d) \frac{x}{d} - \sum_{d \leq x} \mu(d) \left\{ \frac{x}{d} \right\} \tag{1.4}$$

$$1 \geq x \left| \sum_{d \leq x} \frac{\mu(d)}{d} \right| - x + 1$$

Therefore the right term evaluates to:

$$\left| -\sum_{n \leq x} \mu(n) \log(n) \right| \leq x \left| \sum_{d \leq x} \frac{\mu(d)}{d} \right| + x \cdot o(1) \sum_{d \leq x} \frac{1}{d} \leq x + o(x \log(x)) = o(x \log(x))$$

Obtain the final desired relation through summation by parts:

$$\begin{aligned} o(x\log(x)) = \left|\sum_{n\le x}\mu(n)\log(n)\right| &= \left|\log(x)\left(\sum_{n\le x}\mu(n)\right) - \sum_{n\le x}\left(\sum_{d\le n}\mu(d)\right)\log\frac{n+1}{n}\right| \\ &\ge \log(x)\left|\sum_{n\le x}\mu(n)\right| - \sum_{n\le d}\left|\sum_{d\le n}\mu(d)\right|\left(\frac{n+1}{n}-1\right) \\ &\ge \log(x)\left|\sum_{n\le x}\mu(n)\right| - \sum_{n\le x} n\cdot\frac{1}{n} \\ o(x) &\ge \left|\sum_{n\le x}\mu(n)\right| \end{aligned}$$

1.2.3 Assuming the orthogonality

We will now prove the other implication:

$$\boxed{\sum_{n\le x}\mu(n) = o(x) \implies \Psi(x) = (1+o(1))x}$$

Introduce the following identities, that come directly from the Mobius inversion formula:

$$\log(n) = \sum_{d|n}\Lambda(d) \implies \Lambda(n) = \sum_{d|n}\mu(d)\log(\frac{n}{d})$$
$$\tau(n) = \sum_{d|n}1 \implies 1 = \sum_{d|n}\mu(d)\tau(\frac{n}{d})$$

Expand $\Psi - x = \sum_{n\le x}(\Lambda(n)-1)$ using these identities:

$$\begin{aligned} \sum_{n\le x}\Lambda(n) - 1 &= \sum_{n\le x}\sum_{d|n}\mu(d)(\log\frac{n}{d} - \tau(\frac{n}{d})) \\ &= \sum_{m,n:mn\le x}\mu(n)(\tau(m)-\log(m)) \\ &= \sum_{n\le x}\mu(n)\left(\sum_{m\le x/n}\tau(m) - \sum_{m\le x/n}\log(m)\right) \end{aligned}$$

Estimate the two sums. The latter sum is relatively straightforward, through Stirling's formula:

$$\sum_{m\le x/d}\log(m) = \log\left(\frac{x}{d}\right)! = \frac{x}{d}\log\left(\frac{x}{d}\right) - \frac{x}{d} + O\left(\log\left(\frac{x}{d}\right)\right) \tag{1.5}$$

The former can be estimated through Dirichlet's hyperbola method. This is described in many classical

number theory books, for example in [6], page 22. Use $x = x/d$ for easier notation:

$$\begin{aligned}\sum_{m\le x}\tau(m) = \sum_{m\le x}\sum_{ab=m} 1\cdot 1 &= \sum_{a\le\sqrt{x}}\sum_{b\le x/a} 1 + \sum_{b\le\sqrt{x}}\sum_{a\le x/b} 1 - \sum_{a\le\sqrt{x}}\sum_{b\le\sqrt{x}} 1 \\ &= \sum_{a\le\sqrt{x}}\left(\frac{x}{a}+O(1)\right) + \sum_{b\le\sqrt{x}}\left(\frac{x}{b}+O(1)\right) - \left(\sum_{a\le\sqrt{x}} 1\right)^2 \\ &= 2x\sum_{n\le\sqrt{x}}\frac{1}{n} + O(\sqrt{x}) - (\sqrt{x}+O(1))^2 \\ &= 2x\left(\log(\sqrt{x})+\gamma+O\left(\frac{1}{\sqrt{x}}\right)\right) - x + O(\sqrt{x}) \\ &= x\log x + (2\gamma-1)x + O(\sqrt{x})\end{aligned}$$

and so:

$$\sum_{m\le x}\tau(m) = x\log x + (2\gamma-1)x + O(\sqrt{x}) \tag{1.6}$$

Therefore, from (1.5) and (1.6) we can use the following in estimating $\Psi - x$:

$$\sum_{m\le x/n}\tau(m) - \sum_{m\le x/n}\log(m) = 2\gamma\frac{x}{m} + O\left(\sqrt{\frac{x}{m}}\right)$$

$$\begin{aligned}\Psi(x) - x = \sum_{n\le x}\Lambda(n) - 1 &= \sum_{n\le x}\mu(n)\left(2\gamma\frac{x}{n} + O\left(\sqrt{\frac{x}{n}}\right)\right) \\ &= 2\gamma x\sum_{n\le x}\frac{\mu(n)}{n} + \sum_{n\le x} O\left(\sqrt{\frac{x}{n}}\right)\mu(n)\end{aligned}$$

Estimate the two sums, in order to conclude that $\Psi(x) - x = o(x)$. For the first one, use relation (1.4):

$$1 = x\sum_{d\le x}\frac{\mu(d)}{d} - \sum_{d\le x}\mu(d)\left\{\frac{x}{d}\right\}$$

Prove that $\sum_{d\le x}\mu(d)\left\{\frac{x}{d}\right\} = o(x)$. Let $A > 0$, and cut the sum off before applying summation by parts. Use the estimation $\sum_{d\le x}\mu(d) = o(x)$:

$$\begin{aligned}\sum_{x/A\le d\le x}\mu(d)\left\{\frac{x}{d}\right\} &= \sum_{d\le x}\mu(d) + \sum_{n\le x}\left(\sum_{d\le n}\mu(d)\right)\left(\left\{\frac{x}{n+1}\right\} - \left\{\frac{x}{n}\right\}\right) \\ &= o(x) + \sum_{1\le i\le A}\left(\sum_{x/(i+1)\le n<x/i}\left(\sum_{d\le n}\mu(d)\right)\left(\left\{\frac{x}{n+1}\right\} - \left\{\frac{x}{n}\right\}\right)\right) \\ &= o(x) + \sum_{1\le i\le A}\left(o\left(\frac{x}{i}\right)\cdot O(1)\right) = o(x)\end{aligned}$$

so, therefore:

$$x\sum_{d\le x}\frac{\mu(d)}{d} = 1 + \sum_{d\le x}\mu(d)\left\{\frac{x}{d}\right\} \implies \boxed{\sum_{d\le x}\frac{\mu(d)}{d} = o(1)}$$

For the second sum, use the standard integral approximation that shows $\sum_{n\leq x}\frac{1}{n^{1/2}}=O(x^{1/2})$, adapted to show that $\sum_{n\leq x}O\left(\frac{1}{n^{1/2}}\right)=O(x^{1/2})$:

$$\sum_{n\leq x}O\left(\sqrt{\frac{x}{n}}\right)\mu(n)=\sqrt{x}\sum_{n\leq x}O\left(\frac{1}{\sqrt{n}}\right)=O(x)$$

To put everything together, it is required to split the sum in two parts. For simplicity, denote $F\left(\frac{x}{n}\right)=\sum_{m\leq x/n}\tau(m)-\log(m)$ and split the sum according to a variable parameter $1\leq A\leq x$:

$$\begin{aligned}\sum_{n\leq x}\Lambda(n)-1&=\sum_{n\leq x}\mu(n)\left(\sum_{m\leq x/n}\tau(m)-\sum_{m\leq x/n}\log(m)\right)\\&=\sum_{n\leq x/A}\mu(n)F\left(\frac{x}{n}\right)+\sum_{n\geq x/A}^{x}\mu(n)F\left(\frac{x}{n}\right)\\&=\left[2\gamma\frac{x}{A}\sum_{n\leq x/A}\frac{\mu(n)}{n}+\sum_{n\leq x/A}O\left(\sqrt{\frac{x}{n}}\right)\mu(n)\right]+\sum_{k=1}^{A}F(k)\sum_{n=\frac{x}{k+1}}^{\frac{x}{k}}\mu(n)\\&=O\left(\frac{x}{A}\right)+\sum_{k=1}^{A}F(k)o(x)=O\left(\frac{x}{A}\right)+Ao(x)\end{aligned}$$

Since this is true for any A, we can let $x\to\infty$ and $A\to\infty$ such that $Ao(x)=o(x)$ and $O(x/A)=o(x)$ at the same time.

Chapter 2

AC^0 Bounded Depth Circuits.

The purpose of this chapter is to set up the theoretical background for basic circuit complexity theory, and to understand what is a *bounded depth circuit*. In [3], Ben Green proves that the Sarnak conjecture holds for functions $F : \mathbb{N} \to \{-1, 1\}$ such that $F(x)$ is computed from the binary digits of x by a AC^0 bounded depth circuit. We shall define AC^0 and similar other circuit classes, according to standard circuit complexity theory.

This chapter aims to provide a background on circuits and circuit complexity theory, and to prove an important lemma used to study them. In the next chapter, we will prove the Sarnak for AC^0 circuits.

2.1 Boolean Circuits.

2.1.1 Definitions

Definition 2.1 A *boolean circuit* is a directed acyclic graph $C = (V, E)$, where the nodes are partitioned in 3 types: *inputs, gates* and *outputs*. Each node has a set of edges that are incoming, these form the *fan-in* of the node; similarly, the outgoing edges form the *fan-out*.

Denote the number of nodes $|V| = n + s + r$, such that there are n inputs, s gates and r outputs. Then the set of input nodes $X = \{x_1, x_2, \ldots, x_n\}$ is regarded as a set of boolean input variables; each of these nodes has a fan-in of size 0. The set of output nodes $Y = \{y_1, y_2, \ldots, y_r\}$ is regarded as the output of the boolean circuit; each of the nodes has fan-in 1 and fan-out 0.

Now, we want this circuit to be able to compute a multi-variate boolean function $F : \mathbb{Z}_2^n \to \mathbb{Z}_2^r$. To do this, equip each gate $g \in V$ with a boolean function $f : \mathbb{Z}_2^{i(g)} \to \mathbb{Z}_2$, where $i(g), o(g)$ are the sizes of its fan-in and fan-out, respectively. These functions take as inputs the values of the nodes from the fan-in of g, compute a result and distribute it further, to the fan-out of g. In practice, these functions are either one of the standard logical gates AND, OR, NOT.

This is a step-by-step computation, until all gates have an assigned value. Then, the value of each output is equal to the value of the gate from its fan-in (there is only one).

Claim 2.2 *The process does not get stuck.*

Indeed, suppose it did, and there was a configuration in which there was no gate for which all inputs are known. Pick a gate g_0; a gate from its fan-in g_1 has an unknown value. Do the same for g_1, pick a g_2 from its fan-in that is not computed. Continue:

$$g_0 \leftarrow g_1 \leftarrow g_2 \leftarrow \ldots$$

Since the number of gates with unknown inputs is finite, this infinite chain will close at some point, thus yielding a cycle in a DAG, a contradiction.

□

Define the *size* of a circuit as the number of gates it contains. Also define the *depth* of a circuit as the longest path from an input to an output. These concepts will be useful in characterizing circuit families, which we will define now.

A boolean circuit has a set number of inputs, but suppose one wishes to compute a function over inputs of any cardinality:

$$F : Z = \bigcup_{n \geq 1} \mathbb{Z}_2^n \to \mathbb{Z}_2$$

Definition 2.3 Define a *family of circuits* $(C_n)_{n \geq 1}$ such that C_n has n inputs and computes F on the subset $\mathbb{Z}_2^n$. One can analyze asymptotic behaviors of the sizes and widths of C_n as n goes to infinity, and characterize circuit families based on them.

Another important concept in classical circuit complexity theory is *computability*. Intuitively, we want to restrict the families of circuits to only contain circuits that can be computed within a limited amount of time, by a Turing machine, so that the circuits can actually be built fairly efficiently.

For example, a circuit family $(C_n)_{n \geq 1}$ is *logspace uniform* if the following functions are all computable within $O(\log n)$ space. Note that the cumulative information of all these functions yield $(C_n)_{n \geq 1}$ precisely:

- SIZE(n), the function that returns m the size of C_n.
- TYPE$(n, k), 1 \leq k \leq m$, the function that returns the type of the node i. In other words, it returns either AND, OR, NOT, OUTPUT(i), INPUT(j), depending on whether node k is a gate, the ith output or the jth input respectively.
- EDGE(n, i, j) that returns 1 if there is an edge from i to j and 0 otherwise.

Changing the required space to compute these functions yield different uniformity classes; for example, a *P-uniform* family can be computed by functions in $O(n^k)$ space for some k. But, for our purposes, we will use logspace uniformity to define the circuit classes.

2.1.2 Classification. CNF, DNF forms.

Now we will define the main classification classes of boolean circuits, as defined in the previous section. Intuitively, we are interested to define families of circuits that have limited depth, polynomial size and are also logspace uniform.

Definition 2.4 We define the class of $NC^i, i \in \mathbb{N}$ as the class of families of circuits $(C_n)_{n \geq 1}$ that respects the following conditions:

- each gate has at most two inputs, and computes a standard function among AND, OR, NOT;
- has depth $O(\log^i n)$;
- has size $O(n^k)$ for some $k \in \mathbb{N}$;
- is logspace-uniform as defined above.

The class is named after Nick Pippenger (Nick's Class) by Steven Cook, in honor of the researcher of complexity classes.

The AC^i class is similarly defined after the NC^i class, but where all gates are allowed to have an unlimited number of inputs, instead of simply two, therefore calculating AND, OR for a larger number of inputs.

In a circuit C belonging to a family from AC^i, every gate is either a AND or a OR gate, with inputs that are various such other gates. Of course, a AND gate that has another AND gate as an input can be merged in the large one, and the same holds for OR.

In studying this local structure, it makes sense to focus our attention on ANDs of OR gates and vice-versa, on ORs gates of AND. Call the former type a CNF (conjunctive normal form) and the latter ones DNF (disjunctive normal forms).

Definition 2.5 Formally, a CNF on $x_1, x_2, \ldots x_n$ is an AND of ORs such that:

$$f = T_1 \wedge T_2 \wedge \cdots \wedge T_k, \text{where } x_{i1} \vee \cdots \vee x_{in(i)}$$

and a DNF is an OR of ANDs, clearly, the symmetrical:

$$f = T_1 \vee T_2 \vee \cdots \vee T_k, \text{where } T_i = x_{i1} \vee \cdots \vee x_{in(i)}$$

Define the *width* of a CNF/DNF as the maximum size of its components T_i.

Notice that, locally, a circuit in AC^0 looks like a CNF or a DNF. The main idea behind this entire setup is to show that there is significant cancellation within a circuit, and such an evaluation starts from analyzing the behavior of the local CNF/DNF structure. For this purpose, the next section will introduce and prove Hastad's lemma, as defined by Hastad in [7] and proved in a simpler manner by Razborov in [8].

2.2 Hastad's Switching Lemma

The lemma, presented by Johan Hastad in [7], switches a DNF into a CNF and vice-versa, in order to do that in large circuits and collapse their depth at the expense of size. To do so, we connect both of these concepts with a third, common one. Define a *decision tree* in the ordinary way, where at each step the tree branches out with regards to a "decision", that is, a boolean operation with one of the inputs. Naturally, one can associate a decision tree (DT) with a boolean function. Also, similarly to circuits, one can define the *depth* of a DT as being the longest path from its root to one of its leaves.

Thus, CNFs and DNFs can be computed by decision trees. Denote $\mathrm{DT}_{\mathrm{depth}}(f)$ as the depth of the *smallest* circuit that computes the boolean function f.

Hastad's lemma deals with the depth of DNFs under *restrictions*. A restriction ρ represents a mapping from the input variables $x_1, \ldots, x_n$ to the set $\{0, 1, *\}$. A restriction ρ applied to a function f yields a function f_ρ on the variables that are sent to $*$, with the rest fixed to either 0 or 1.

A *random restriction* ρ with parameter $0 \leq p \leq 1$ is a restriction where $\rho(x_i)$ is randomly set:

$$\rho(x_i) = \begin{cases} * & \text{with probability } p \\ 0 & \text{with probability } \frac{1-p}{2} \\ 1 & \text{with probability } \frac{1-p}{2} \end{cases}$$

Intuitively, we want to prove that a DNF suffers significant reductions under a restriction, in most cases. The formal statement of the lemma:

Theorem 2.6 *Hastad's switching lemma. Let f be a DNF of width w and n variables, and ρ a random restriction with parameter $p = \sigma n, \sigma \leq 1/5$. Then, for each parameter d:*

$$Pr\left(DT_{depth}(f_\rho) > d\right) \leq (10\sigma w)^d$$

We wish to show that there are relatively few **bad restrictions** ρ such that f_ρ requires a decision tree with depth greater than d to be computed. In other words, if $\mathcal{B}$ is the set of **bad restrictions** and $\mathcal{R}_p$ is the set of **all restrictions** with cardinality p, then we wish to show that:

$$\frac{|\mathcal{B}|}{|\mathcal{R}_p|} \leq (10\sigma w)^d$$

The cardinality of $\mathcal{R}_p$ is easily computable; it is $\binom{n}{p}2^{n-p}$. The main idea is to reduce the restriction $\rho \in \mathcal{R}_p$ to another restriction $\rho' \in \mathcal{R}_{p-d}$, plus some "additional information" that would ensure injectivity. This set of "additional information" shall be defined later.

Definition 2.7 We shall now define, for a DNF f, its *canonical decision tree*. Such a tree will, in particular, have depth greater than $\mathrm{DT}_{\mathrm{depth}}(f)$. If $T_1, T_2, \ldots, T_k$ are the terms of f, then take the first term T_1, say it has length d_1, and order its literals. Construct a decision tree of depth d_1 that passes through all the literals of T_1 in order. T_1 is an AND; therefore, exactly one of the end-paths corresponds to 1, where f is also 1. For the rest of the paths, continue calculating T_2 similarly while omitting the already decided variables, and so on until the end.

A similar procedure is applied for f_ρ. Under ρ, many of the terms T_i will cancel out; it is enough for a single literal in T_i to be set to 0, and the entire term is trivialized. Suppose that ρ does not kill the terms $T_{i_1}, \ldots, T_{i_k}$, and denote the non-trivial reductions under ρ of the terms T_i as $U_j, 1 \leq j \leq k'$. Therefore:

$$f_\rho = U_1 \vee U_2 \vee \cdots \vee U_{k'}$$

and one can build the canonical decision tree for this function. By hypothesis, the depth of this tree is greater than d.

Definition 2.8 Now, define π as the lexicographical leftmost path of length greater than d, and consider only its first d terms; that is, the first d variables that π restricts. Because the decision tree is still not trivial, this means that, under the restriction $\rho\pi \in \mathcal{R}_{p-d}$ which is the combined restriction of ρ and π, the function $f_{\rho\pi}$ is not decided. Unfortunately, using this new restriction will not be enough to recover ρ from $\rho\pi$, since it is difficult to see what belongs to this path and what not.

Instead, let us look at $U_1, U_2, \ldots U_{k'}$. Denote γ_i as the (unique) restriction that fixes U_i to 1, and π_i the part of π that fixes variables from γ_i. Assuming that π_1 is not the entirety of π, note that $\beta\pi_1$ kills the term U_1. Otherwise, since U_1 is an AND, all its literals fixed by π_1 would also be fixed by γ_1 in the same way, and at the end of π_1 the term U_1 would be undecided, which is impossible by definition; π_1 ends and π_2 begins when U_1 is decided.

Continue this process and define γ_j, π_j for each T_{i_j}, U_j. The restriction that we are looking for is:

$$\rho' = \rho\gamma_1\gamma_2 \ldots \gamma_k$$

Note that, because γ_j fixes the same number of variables as π_j, $\rho' \in \mathcal{R}_{p-d}$, just as $\rho\pi$. Now we just have to decide on the additional information that is to be added, so that ρ can be recovered from ρ'.

Claim 2.9 The function $f_{\rho'}$ fixes $T_1, T_2, \ldots T_{i_1-1}$ to 0, exactly as f_ρ does. Also, T_{i_1} is the first term that $f_{\rho'}$ fixes to 1.

The first part of the claim is easy to establish, since $\rho \subset \rho'$ and T_{i_1}, by definition, is the first term not to be fixed to 0. Also, by the definition of γ_1, U_1 is fixed to 1 and since T_{i_1} was undecided, now it is also fixed to 1.

After identifying i_1, we want to identify which of the variables in T_{i_1} are fixed by γ_1 and which do not. If $|U_1| = u_1$, then this information will require $u_1 \log w$ bits of information, since $|T_{i_1}| \leq w$ and any position from 1 to w can be encoded in $\log w$ bits of information. Also encode how π_1 fixes these variables with u_1 additional bits. Together with the encoding of u_1 itself, needed to revert the operation, this amount of information occupies a certain number of bits:

$$u_1 \log w + u_1 + \log u_1 \leq u_1 \log w + 2u_1$$

Now, based on this information, we can retrieve γ_1 from the knowledge of its variables and π_1 as well, from the knowledge of its assignments. For the next step, consider $\rho'_1 = \rho\pi_1\gamma_2 \ldots \gamma_k$. Following the exact same argument, the first term of f that this restriction fixes to 1 is T_{i_2}., so the same procedure can be applied. In $u_2 \log w + 2u_2$ bits, one can find out γ_2, π_2. Continue similarly until the end of the restriction π, when we determine the last γ_k. In total, all the γ_j should fix precisely d variables.

At this point, we have determined $\gamma_1, \ldots, \gamma_k$ which means that we can recover ρ from ρ'. This construction is in fact an injection:

$$\mathcal{R}_p \hookrightarrow \mathcal{R}_{p-d} \times S$$

where S stands for the set of binary strings that encode the additional information needed to establish the reverse operation, as described. S has a total size of:

$$(u_1 \log w + 2u_1) + (u_2 \log w + 2u_2) + \cdots + (u_k \log w + 2u_k) = d \log w + 2d$$

Claim 2.10 The size of the term on the right side is small enough so that the proportion of "bad" restrictions is small as well, as Hastad's lemma states.

This requires a final computation of the term's size:

$$\begin{aligned} |\mathcal{R}_{p-d} \times S| &= \left|\mathcal{R}_{p-d} \times \{0,1\}^{d \log w + 2d}\right| \\ &= \binom{n}{p-d} 2^{n-(p-d)} (4w)^d \end{aligned}$$

and therefore the ratio of "bad" restrictions is:

$$\begin{aligned} \frac{|\mathcal{R}_{p-d} \times S|}{|\mathcal{R}_p|} &= \frac{\binom{n}{p-d} 2^{n-(p-d)} (4w)^d}{\binom{n}{p} 2^{n-p}} \\ &= (8w)^d \frac{p(p-1)\ldots(p-d+1)}{(n-p+d)(n-p+d-1)\ldots(n-p+1)} \\ &\leq (8w)^d \left(\frac{p}{n-p+d}\right)^d \\ &\leq (8w)^d \left(\frac{\sigma}{1-\sigma}\right)^d \leq (10\sigma w)^d \text{ , because } \sigma \leq \frac{1}{5}. \end{aligned}$$

thereby proving the Hastad lemma 2.6. □

Claim 2.11 If f is a Boolean function that has:

$$\mathrm{DT}_{\mathrm{depth}}(f) \leq d$$

then there is a DNF of depth d that computes f; namely, the OR of all 1-leaves of the tree.

In order to analyze circuits properly, we will however need a symmetrical version of Hastad's lemma. This can be established by noticing the natural bijection between CNFs and DNFs. Formally, if f is a CNF (that is, an AND of ORs):

$$f = T_1 \wedge T_2 \wedge \cdots \wedge T_k, \text{where } T_i = x_{i1} \vee \cdots \vee x_{in(i)}$$

then the negation $\neg f$ is, by De Morgan's law:

$$\begin{aligned} \neg f &= \neg\left(T_1 \wedge T_2 \wedge \cdots \wedge T_k\right) \\ &= \neg T_1 \vee \neg T_2 \vee \cdots \vee \neg T_k = \bigvee_{i=1}^{k} \neg T_i \\ &= \bigvee_{i=1}^{k} \neg\left(x_{i1} \vee \cdots \vee x_{in(i)}\right) \\ &= \bigvee_{i=1}^{k} \bigwedge_{j=1}^{n(i)} \overline{x_{ij}} \end{aligned}$$

in other words, a DNF. Because f and $\neg f$ are computational-wise analogous, this essentially proves that:

Claim 2.12 Hastad's lemma 2.6, and its "reciprocal" 2.11 provides the same relation for CNFs and DNFs.

Chapter 3

Sarnak Conjecture for AC^0 Circuits

This chapter studies the conjecture 1.1 applied to functions determined by bounded depth circuits in AC^0. It relies on two main results, one of which will be proven here and another one in the next chapter.

3.1 The main argument.

Let $N = 2^n$ and $F : \{0, 1, \ldots, N-1\} \to \{-1, 1\}$ be a bounded depth circuit function in $AC^0(d)$. This section will prove the following estimation, that implies the desired result:

$$\frac{1}{N} \sum_{x<N} \mu(x)F(x) = O\left(e^{d \log n - cn^{\frac{1}{6}d}}\right)$$

for $c > 0$ a constant.

We first introduce the Fourier-Walsh coefficients, that play a critical role in this proof.

Definition 3.1 For any $x \in \{0, 1, \ldots N-1\}$, consider its expansion in binary digits $x = x_0 + 2x_1 + \ldots 2^{n-1}x_{n-1}$. For a function $F : \{0, 1, \ldots, N-1\} \to \{-1, 1\}, F \in AC^0(d)$ and for $S \subset \{0, 1, \ldots n-1\}$ define the *Fourier-Walsh coefficient* of F:

$$\hat{F}(S) = \frac{1}{N} \sum_{x<N} F(x)(-1)^{\sum_{i \in S} x_i}$$

For shortness, define $s_S(x) = \sum_{i \in S} x_i$, and the *character* $\chi_S(x) = (-1)^{s_S(x)}$.

Green's proof revolves around two main lemmas, the combination of which yields the main result rather quickly.

Theorem 3.2 First, an important result of Linial, Mansour and Nisan from [5] gives us an upper bound on the size of the Fourier-Walsh coefficients for a function F computed by a bounded depth circuit of depth d and size M:

$$\boxed{\sum_{|S|>t} |\hat{F}(S)|^2 \leq 2M2^{-t^{1/d}/40}}$$

Theorem 3.3 Second, the original content of the paper, is a bound on the size of Fourier-Walsh coefficients for the Mobius function. If $S \subset \{0, 1, \ldots n-1\}$, $|S| = k$ then:

$$\boxed{\hat{\mu}(S) = O(ke^{-cn^{1/2}/k})}$$

for $c > 0$ a constant.

Let us establish an identity that is a discrete version of Parseval's identity regarding sums of Fourier coefficients. For completeness, we shall prove the discrete version here:

$$\begin{aligned}\sum_S \hat{\mu}(S)\hat{F}(S) &= \frac{1}{N^2}\sum_S \left(\sum_{x<N}\mu(x)(-1)^{s_S(x)}\right)\left(\sum_{x<N}F(x)(-1)^{s_S(x)}\right)\\ &= \frac{1}{N^2}\sum_{x,y<N}\mu(x)F(y)\sum_S(-1)^{s_S(x)+s_S(y)}\\ &= \frac{1}{N}\sum_{x<N}\mu(x)F(x)\end{aligned}$$

We used the following proposition:

Proposition 3.4 The sum $\sum_S(-1)^{s_S(x)+s_S(y)}$ can be calculated as:

$$\sum_S(-1)^{s_S(x)+s_S(y)} = \begin{cases} N & \text{for } x=y \\ 0 & \text{otherwise}\end{cases}$$

If $x=y$, then $s_S(x)=s_S(y)$ and so the sum is clearly $N=2^n$, the number of different subsets. If $x\neq y$, then they differ on a certain position $0\leq k\leq n-1$, assume WLOG that the kth digit of x is 0 and of y is 1.

Pair every subset S for which $k\notin S$ with the subset $S'=S\cup\{k\}$, and notice that this represents a partition of all subsets in pairs, and that $s_{S'}(x)=s_S(x), s_{S'}(y)=s_S(y)+1$. Put everything together:

$$\begin{aligned}\sum_S(-1)^{s_S(x)+s_S(y)} &= \sum_{k\notin S}(-1)^{s_S(x)+s_S(y)}+(-1)^{s_{S'}(x)+s_{S'}(y)}\\ &= \sum_{k\notin S}(-1)^{s_S(x)+s_S(y)}+(-1)^{s_S(x)+s_S(y)+1}\\ &= 0\end{aligned}$$

□

Returning to the main proof, we now established that, for a function $F\in AC^0(d)$, that is, of depth d and size at most n^d:

$$\boxed{\frac{1}{N}\sum_{x<N}\mu(x)F(x)=\sum_S\hat{\mu}(S)\hat{F}(S)}$$

Split the sum in two parts, by comparing $|S|$ and $n^{1/6}$; we want to estimate the modulus of this sum. For $|S|<n^{1/6}$, use the trivial $\hat{F}(S)\leq 1$ and bound the sum by:

$$2^{n^{1/6}}\sup_{|S|<n^{1/6}}|\hat{\mu}(S)|+\sum_{|S|>n^{1/6}}|\hat{\mu}(S)||\hat{F}(S)|$$

For the first term, assuming the second statement, we get that:

$$2^{n^{1/6}}\sup_{|S|<n^{1/6}}|\hat{\mu}(S)| = 2^{n^{1/6}}O(n^{1/6}e^{-cn^{1/3}}) = O(e^{-cn^{1/3}})$$

For the second term, use the Cauchy-Schwarz inequality:

$$\sum_{|S|>n^{1/6}}|\hat{\mu}(S)||\hat{F}(S)| \leq \left(\sum_{|S|\geq n^{1/6}}|\hat{\mu}(S)|^2\right)^{1/2}\left(\sum_{|S|\geq n^{1/6}}|\hat{F}(S)|^2\right)^{1/2}$$

The first sum can be estimated by using Parseval's identity, in a very similar fashion to the one proved above:

$$\sum_{|S|\geq n^{1/6}} |\hat{\mu}(S)|^2 \leq \sum_{S} \hat{\mu}(S)\cdot\hat{\mu}(S) = \frac{1}{N}\sum_{x<N}\mu(x)^2 \leq 1$$

and the second sum is estimated by assuming the first lemma 3.2:

$$\sum_{|S|\geq n^{1/6}} |\hat{F}(S)|^2 \leq 2n^d 2^{-c'n^{1/6d}}$$

Putting everything together, we obtain the following upper bound:

$$\begin{aligned}\frac{1}{N}\sum_{x<N}\mu(x)F(x) &\leq O(e^{-cn^{1/3}}) + \left(2n^d 2^{-c'n^{1/6d}}\right)^{1/2} \\ &\leq O\left(e^{d\log n - c'n^{1/6d}}\right) \\ &\leq o(1)\end{aligned}$$

as required, therefore proving the main result.

3.2 Linial, Mansour and Nisan result

An important tool in proving **Theorem 3.2** will be Hastad's **Lemma 2.6**, already discussed in the previous chapter. We restate it here.

Hastad's Lemma. Let f be a DNF (or a CNF, by proposition 2.12) of width w and n variables, and ρ a random restriction with parameter $p = \sigma n, \sigma \leq 1/5$. Then, for each parameter s:

$$Pr\left(\mathrm{DT}_{\mathrm{depth}}(f_\rho) > s\right) \leq (10\sigma w)^s$$

and therefore, by the fact 2.11, we can write it as a CNF (or a DNF, respectively) with width smaller than s with a probability of at least 1 - $(10\sigma w)^s$. Furthermore, all clauses of this DNF have disjoint inputs.

Define the *degree* of a Boolean function as the size of the largest subset that does not kill f:

$$\deg(f) = \max_{\hat{f}(S)\neq 0} |S|$$

and prove the following corollary of Hastad's lemma regarding the degree of a Boolean function.

Corollary 3.5 For f a Boolean function, the following holds:

$$\mathrm{DT}_{\mathrm{depth}}(f) \geq \deg(f)$$

This takes a bit of insight to notice. Assume that this is not the case; that is, if $\mathrm{DT}_{\mathrm{depth}}(f) = d$ there exists a subset $S, |S| > d$ such that $\hat{f}(S) \neq 0$. Look at the decision tree of f, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the restrictions that they impose on the set of variables such that:

$$\lambda_1 \sqcup \lambda_2 \sqcup \cdots \sqcup \lambda_k = \{0,1\}^n$$

Because $\mathrm{DT}_{\mathrm{depth}}(f) = d$, each λ_i fixes at most d variables. And because $|S| > d$, for each λ_i there exists

a variable $x_{n(i)} \in S, \notin \lambda_i$. Now, we can assert that $\hat{f}(S) = 0$:

$$\begin{aligned}\hat{f}(S) &= \frac{1}{N} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x) \\ &= \frac{1}{N} \sum_{i=1}^{k} \sum_{x \in \lambda_i} f(x)\chi_S(x) \\ &= \frac{1}{N} \sum_{i=1}^{k} \sum_{x \in \lambda_i, x_{n(i)}=0} f(x)\chi_S(x)(1-1) = 0\end{aligned}$$

thus proving that $\mathrm{DT}_{\mathrm{depth}}(f) \geq \deg(f)$.

□

We shall use Hastad's lemma repeatedly to prove the following lemma about circuits.

Lemma 3.6 Let f be a Boolean function computed by a circuit of size M and depth d. Then, if ρ is a random restriction with parameter:

$$Pr\left(\rho(x_i) = *\right) = \frac{1}{20^d s^{d-1}}$$

for some s, then:

$$Pr\left(\mathrm{DT}_{\mathrm{depth}}(f_\rho) > s\right) \leq M 2^{-s}$$

View the circuit of f as stratified on d successive levels, each level with m_i gates. Naturally, $\sum m_i = M$. Also, view the random restriction ρ as a succession of random restrictions, the first one with parameter $1/20$, and $d-1$ successive ones with parameter $1/20s$.

After the first restriction, with high probability, most of the gates from the lowest level will have at most s inputs; of course, this step is needed to ensure the application of Hastad's lemma to the further steps. To see this, for each gate on this level, consider two cases:

- If the original number of inputs is at least $2s$, then the chances that this reduction kills it, that is, sets one of its inputs to 0 and is a AND or to 1 and is a OR, is at most:

$$0.525^{2s} \leq 2^{-s}$$

- If the original number of inputs is at most $2s$, then the probability that at least s inputs get assigned a $*$ is:

$$\binom{2s}{s} 20^{-s} \leq 2^{-s}$$

Therefore, because there are m_1 gates at this level, the probability of a failure is **at most** $m_1 2^{-s}$. For each successive step, the idea is to convert all CNFs into DNFs (or vice-versa) for the second and third layers from the bottom, so that they can be collapsed into a single level and thus reducing the depth of the entire circuit.

By induction, we know that these gates composing the CNF (or DNF) have less than s inputs; or, in other words, width $\leq s$. Therefore, when we switch it into a DNF (or CNF) under a random restriction with parameter $p = 1/20s$, by using Hastad's switching lemma we know its probability to have width more than s:

$$Pr\left(\mathrm{DT}_{\mathrm{depth}}(f_\rho) > s\right) \leq \left(10 \frac{1}{20s} s\right)^s = 2^{-s}$$

This kind of switches lets us collapse the second and third levels above the input levels together, thus reducing the depth of the circuit by 1. The probability that this fails is at most $m_i 2^{-s}$, since for each gate the probability is at most 2^{-s}.

After $d-2$ switches, we are left with a CNF (or DNF) that has width at most s. By finally using Hastad's lemma again for the last reduction, we conclude that $\mathrm{DT}_{\mathrm{depth}}(f_\rho) \leq 2^{-2}$. Adding up all the probabilities yields what we are looking for:

$$Pr(\mathrm{DT}_{\mathrm{depth}}(f_\rho) > s) \leq M2^{-s}$$

□

Now, we start investigating the Fourier-Walsh coefficients, as defined at the beginning of the chapter in 3.1. Remember the definition, for $N = 2^n$ and a subset $S \subset \overline{1,n}$:

$$\hat{f}(S) = \frac{1}{N} \sum_{x \leq N} f(x)\chi_S(x)$$

where $\chi(S) = (-1)^{\sum x_i}$ and x_i are the binary digits of x. One can view this sum as running over $\{0,1\}^n$ the vertices of the n-dimensional cube. The following few lemmas will see how the Fourier-Walsh coefficients break down under a restriction, in order to make a connection with circuit restrictions.

Lemma 3.7 Let $S \subset \overline{1,n}$ be a subset of the variables. Then, for any other subset A the following relation holds:

$$\hat{f}(A) = 2^{-|\bar{S}|} \sum_{x \in \{0,1\}^{|\bar{S}|}} \chi_{A \cap \bar{S}}(x) \hat{f}_{\bar{S} \leftarrow x}(A \cap S)$$

where $\hat{f}_{\bar{S} \leftarrow x}$ is the restriction of $\hat{f}$ to only the vertices in S, and on the vertices of $\bar{S}$ the values of x are used. In other words, if $R \in \{0,1\}^n$ such that $R|_S = x$ and $R|_{\bar{S}} = y$, then:

$$f_{\bar{S} \leftarrow x}(y) = f(R)$$

The proof of this lemma is fairly straightforward. Begin by unpacking the term on the right, using the definition of $\hat{f}$:

$$\begin{aligned}
RHS &= 2^{-|\bar{S}|} \sum_{x_1 \in \{0,1\}^{|\bar{S}|}} \chi_{A \cap \bar{S}}(x_1) 2^{-|S|} \sum_{x_2 \in \{0,1\}^{|S|}} \chi_{A \cap S}(x_2) f_{\bar{S} \leftarrow x_1}(x_2) \\
&= 2^{-|S|-|\bar{S}|} \sum_{x_1 \in \{0,1\}^{|\bar{S}|}} \sum_{x_2 \in \{0,1\}^{|S|}} \chi_{A \cap \bar{S}}(x_1) \chi_{A \cap S}(x_2) f_{\bar{S} \leftarrow x_1}(x_2) \\
&= 2^{-n} \sum_{x \in \{0,1\}^n} \chi_A(x) f(x) \\
&= \hat{f}(A)
\end{aligned}$$

where $x \in \{0,1\}^n$ such that $x|_{\bar{S}} = x_1$ and $x|_S = x_2$. Under this notation, it becomes straightforward that $\chi_{A \cap \bar{S}}(x_1)\chi_{A \cap S}(x_2) = \chi_A(x)$, since one is just adding binary digits, and that $f_{\bar{S} \leftarrow x_1}(x_2) = f(x)$ by definition. □

Lemma 3.8 Let f be a Boolean function and $S \in \overline{1,n}$ an arbitrary subset of variables. Then, for any subset $B \subset S$, we have:

$$\sum_{C \subset \bar{S}} \hat{f}(B \cup C)^2 = 2^{-|\bar{S}|} \sum_{x \in \{0,1\}^{|S|}} \hat{f}_{\bar{S} \leftarrow x}(B)^2$$

Apply 3.7 to expand the term on the left hand side, keeping in mind that $B \subset S$ and $C \subset \bar{S}$:

$$\begin{aligned}
\sum_{C\subset\bar{S}} \hat{f}(B\cup C)^2 &= \sum_{C\subset\bar{S}} \left(2^{-|\bar{S}|} \sum_{x\in\{0,1\}^{|\bar{S}|}} \chi_C(x) \hat{f}_{\bar{S}\leftarrow x}(B) \right)^2 \\
&= 2^{-2|\bar{S}|} \sum_{C\subset\bar{S}} \left(\sum_{x_1\in\{0,1\}^{|\bar{S}|}} \sum_{x_2\in\{0,1\}^{|\bar{S}|}} \chi_C(x_1)\chi_C(x_2) \hat{f}_{\bar{S}\leftarrow x_1}(B) \hat{f}_{\bar{S}\leftarrow x_2}(B) \right) \\
&= 2^{-|\bar{S}|} \sum_{x_1,x_2\in\{0,1\}^{|\bar{S}|}} \hat{f}_{\bar{S}\leftarrow x_1}(B) \hat{f}_{\bar{S}\leftarrow x_2}(B) \left[2^{-|\bar{S}|} \sum_{C\subset\bar{S}} \chi_C(x_1 \oplus x_2) \right]
\end{aligned}$$

By referring to Proposition 3.4 and applying it to the sum inside the square brackets, one concludes that the sum is 1 if $x = y$ and 0 otherwise. Therefore, the expression becomes exactly what we are looking for:

$$\sum_{C\subset\bar{S}} \hat{f}(B\cup C)^2 = 2^{-|\bar{S}|} \sum_{x\in\{0,1\}^{|S|}} \hat{f}_{\bar{S}\leftarrow x}(B)^2$$

□

Lemma 3.9 Let f be a Boolean function, $S \subset \overline{1,n}$ a subset of variables and k an integer. Then the following relation holds:

$$\sum_{A,|A\cap S|>k} \hat{f}(A)^2 \leq Pr\left[\deg(f_{\bar{S}\leftarrow R}) > k\right] \leq Pr\left[\mathrm{DT}_{\mathrm{depth}}(f_{\bar{S}\leftarrow R}) > k\right]$$

where R is a random assignment in $\{0,1\}^{\bar{S}}$ for the variables in $\bar{S}$.

The second inequality comes naturally from 3.5. The proof for the first inequality relies on the fact that for any $A, |A \cap S| > \deg(f_{\bar{S}\leftarrow R})$, we have $\hat{f}(A \cap S) = 0$, naturally by the definition of deg. Because $f^{\widehat{\ }}_{\bar{S}\leftarrow R}(B)^2 > 1$, it is enough to show that:

$$\sum_{A,|A\cap S|} \hat{f}(A)^2 \leq \mathbb{E}_R \left[\sum_{|B|>k} \hat{f}_{\bar{S}\leftarrow R}(B)^2 \right]$$

because the term in parenthesis is a nonnegative integer. Let us rewrite the left hand side, apply Lemma 3.8 and switch the sums:

$$\begin{aligned}
\sum_{A,|A\cap S|>k} \hat{f}(A)^2 &= \sum_{B\subset S,|B|>k} \sum_{C\subset\bar{S}} \hat{f}(B\cup C)^2 \\
&= \sum_{B\subset S,|B|>k} 2^{-|\bar{S}|} \sum_{x\in\{0,1\}^{|S|}} \hat{f}_{\bar{S}\leftarrow x}(B)^2 \\
&= 2^{-|\bar{S}|} \sum_{x\in\{0,1\}^{|S|}} \left[\sum_{B\subset S,|B|>k} \hat{f}_{\bar{S}\leftarrow x}(B)^2 \right] \\
&= \mathbb{E}_R \left[\sum_{|B|>k} \hat{f}_{\bar{S}\leftarrow R}(B)^2 \right]
\end{aligned}$$

therefore finishing the proof.

□

The following lemma averages sums from 3.9 over sets S.

Lemma 3.10 Let f be a Boolean function, t an integer and $0 < p < 1$. Then:

$$\sum_{|A|>t} \hat{f}(A)^2 \leq 2\mathbb{E}_S \left[\sum_{|A\cap S|>pt/2} \hat{f}(A)^2 \right]$$

where $S \subset \overline{1,n}$ is randomly chosen, with each element having probability p to belong to S, and $pt > 8$.

Using a Chernoff bound, the probability for $|A \cap S| > pt/2$ is larger than $1 - \exp(-pt/8) > 1/2$. Therefore, each set A contributes with a probability of at least $1/2$ for the sets on the left side, and the lemma is proven.

The proof for the bound used above will be presented now. Let $A \subset \overline{1,n}$ be a fixed subset, $|A| = t$, and $S \subset \overline{1,n}$ be a random subset, where each $x \in \overline{1,n}$ has $Pr(x \in S) = p$ for a parameter $0 < p < 1$. Then:

$$Pr(|A \cap S| > pt/2) > 1 - \exp(-pt/8)$$

Let $A = \{a_i : 1 \leq i \leq n\}$ and $x_i, 1 \leq i \leq n$ be random variables such that $x_i = 1 \iff a_i \in S$, and 0 otherwise. Let $X = \sum x_i$; we want to evaluate $Pr(X > pt/2)$.

The proof relies on applying Markov's inequality to the exponential $e^{\lambda X}$ for $\lambda > 0$. Recall Markov's inequality:

$$Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

and apply it as follows:

$$\begin{aligned} Pr(X \leq pt/2) = Pr(e^{-\lambda X} \geq e^{-\lambda pt/2}) &\leq e^{\lambda pt/2}\mathbb{E}(e^{-\lambda X}) \\ &\leq e^{\lambda pt/2}\mathbb{E}(e^{-\lambda x_i})^t \\ &\leq e^{\lambda pt/2}\left(1 + p(e^{-\lambda} - 1)\right)^t \\ &\leq e^{\lambda pt/2} e^{pt(e^{-\lambda}-1)} \\ &\leq \exp\left(pt\left(\frac{\lambda}{2} + e^{-\lambda} - 1\right)\right) \end{aligned}$$

Therefore, because $\lambda > 0$ is a parameter we can choose, pick for example $\lambda = 1$ and notice that:

$$\frac{\lambda}{2} + e^{-\lambda} - 1 = e^{-1} - \frac{1}{2} = -0.13 \leq -\frac{1}{8}$$

which proves the required relation.

□

We are now ready to prove the main result 3.2. We shall restate it here:

Lemma 3.2 For a function f computed by a bounded depth circuit of depth d and size M, the following holds:

$$\sum_{|S|>t} |\hat{f}(S)|^2 \leq 2M2^{-t^{1/d}/40}$$

Fix $p = 1/(20t^{(d-1)/d})$ and $s = pt/2 = t^{1/d}/40$. By lemma 3.10, we know the following:

$$\sum_{|A|>t} \hat{f}(A)^2 \leq 2\mathbb{E}_S \left[\sum_{|A\cap S|>pt/2} \hat{f}(A)^2 \right]$$

with $S \subset \overline{1,n}$ a random subset, where every element belongs to S with probability p.

Using lemma 3.9, we know the following bound:

$$2\mathbb{E}_S \left[\sum_{|A\cap S|>pt/2} \hat{f}(A)^2 \right] \leq 2\mathbb{E}_S Pr\left[\mathrm{DT}_{\mathrm{depth}}(f_{\bar{S}\leftarrow R}) > \frac{pt}{2} = s \right]$$

The key observation now is that a selection of variables S randomly and uniformly with probability p and then a random association of $\{0,1\}$ values to $\bar{S}$ is exactly the same thing as a random restriction ρ with parameter p. Therefore, lemma 3.6 can be applied, because $p = 1/(20^d s^{d-1})$ and so we can conclude:

$$\sum_{|A|>t} \hat{f}(A)^2 \leq 2M2^{-s} = 2M2^{-t^{1/d}/40}$$

thereby finishing the proof.

Chapter 4

Proof of Ben Green's Relation

In this last chapter we prove the second required **Lemma 3.3**, following the argument presented in [3]. Recall the statement of 3.3: If $S \subset \{0, 1, \dots n-1\}$, $|S| = k$ then:

$$\hat{\mu}(S) = O(ke^{-cn^{1/2}/k})$$

for $c > 0$ a constant.

The proof set forth by Ben Green in [3] begins by establishing a connection between the Fourier-Walsh coefficients as defined above and the "traditional" Fourier coefficients:

$$\hat{F}(\theta) = \frac{1}{N} \sum_{x<N} F(x) e^{2i\pi x\theta},\ \theta \in [0, 1]$$

The first result we will prove is according to Katai.

Theorem 4.1 *(Katai)* Assume there exists a function $F : \{0, 1, \dots N-1\} \to [-1, 1]$ such that there is a subset $S \subset \{0, 1, \dots n-1\}, |S| = k$ and the corresponding Fourier-Walsh coefficient $|\hat{F}(S)| > \delta, 0 < \delta < 1/2$. Then there is a $\theta \in [0, 1]$ such that the Fourier coefficient $|\hat{F}(\theta)| > (\delta/10k)^{4k}$.

Moreover, this θ can be taken to be a *sparse dyadic* rational:

$$\theta = \frac{r_1}{2^{i_1}} + \frac{r_2}{2^{i_2}} + \cdots + \frac{r_k}{2^{i_k}},\ r_i \in \mathbb{Z}, |r_i| \leq (10k/\delta)^3$$

4.1 Proof of Katai's result

Define $\psi : \mathbb{R}/\mathbb{Z} \to [-1, 1]$, an indicator function:

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq t < 1 \end{cases}$$

and notice that we can rewrite $s_S(x)$ mod 2 as follows:

$$s_S(x) = \sum_{i \in S} x_i \equiv \prod_{i \in S} \psi\left(\frac{x}{2^i}\right)$$

because $x_i = 1 \iff \psi(x/2^i) = -1$ and $x_i = 0 \iff \psi(x/2^i) = 1$. Intuitively, both sums count the parity of the number of odd digits of x in binary form. This rewriting also translates into a rewriting of the Fourier-Walsh coefficient:

$$\hat{F}(S) = \frac{1}{N} \sum_{x<N} F(x) \prod_{i \in S} \psi\left(\frac{x}{2^i}\right)$$

The main idea is to replace the function ψ by a function $\tilde{\psi}$ that is both close enough to ψ and has a reasonable Fourier expansion. Namely, we want the following two conditions to be satisfied, and we shall prove that such a function can be constructed through a smoothing procedure:

- Close to ψ, or, more specifically, for any $0 \leq i \leq n-1$:

$$\boxed{\frac{1}{N}\sum_{x<N}\left|\psi\left(\frac{x}{2^i}\right)-\tilde{\psi}\left(\frac{x}{2^i}\right)\right| \leq \epsilon}$$

- With a relatively well-behaved Fourier expansion:

$$\boxed{\tilde{\psi}(t)=\sum_{r\leq 100\epsilon^{-3}} a_r e^{2i\pi rt}}$$

  such that $|a_r| \leq 1$ for all r.

First, we assume the above construction and present the remainder of the proof. Choose $\epsilon = \delta/2k$ and replace ψ with a $\tilde{\psi}$ obtained through the above construction. Starting off with the original assumption, apply the triangle inequality repeatedly to achieve a lower bound for the Fourier coefficient:

$$|\hat{F}(S)| = \frac{1}{N}\left|\sum_{x<N} F(x)\prod_{i\in S}\psi\left(\frac{x}{2^i}\right)\right| \geq \delta$$

$$\frac{1}{N}\left|\sum_{x<N} F(x)\left(\tilde{\psi}\left(\frac{x}{2^{i_1}}\right)+\left(\psi\left(\frac{x}{2^{i_1}}\right)-\tilde{\psi}\left(\frac{x}{2^{i_1}}\right)\right)\right)\prod_{i\in S, i\neq i_1}\psi\left(\frac{x}{2^i}\right)\right| \geq \delta$$

$$\frac{1}{N}\left|\sum_{x<N} F(x)\tilde{\psi}\left(\frac{x}{2^{i_1}}\right)\prod_{i\in S, i\neq i_1}\psi\left(\frac{x}{2^i}\right)\right| - \frac{\epsilon}{N}\left|\sum_{x<N} F(x)\prod_{i\in S, i\neq i_1}\psi\left(\frac{x}{2^i}\right)\right| \geq \delta$$

Since F and ψ only take values between $[-1,1]$, the second sum is at most ϵ. Because $|S| = k$, this procedure of extracting one ψ at a time will eventually yield:

$$\frac{1}{N}\left|\sum_{x<N} F(x)\prod_{i\in S}\tilde{\psi}\left(\frac{x}{2^i}\right)\right| - k\epsilon \geq \delta$$

Expand every ψ in its Fourier series, and obtain:

$$\frac{1}{N}\left|\sum_{x<N} F(x)\prod_{i\in S}\left(\sum_{r_i\leq 100\epsilon^{-3}} a_r e^{2i\pi r_i x/2^i}\right)\right| \geq \delta/2$$

$$\frac{1}{N}\left|\sum_{x<N} F(x)\sum_{r_1\ldots r_k<100\epsilon^{-3}} a_{r_1}\ldots a_{r_k}\exp\left(2i\pi x\sum_{i\in S}\frac{r_i}{2^i}\right)\right| \geq \delta/2$$

$$\left|\sum_{r_1\ldots r_k<100\epsilon^{-3}} a_{r_1}\ldots a_{r_k}\left(\frac{1}{N}\sum_{x<N} F(x)\exp\left(2i\pi x\sum_{i\in S}\frac{r_i}{2^i}\right)\right)\right| \geq \delta/2$$

$$\left|\sum_{r_1\ldots r_k<100\epsilon^{-3}} a_{r_1}\ldots a_{r_k}\tilde{F}\left(\sum_{i\in S}\frac{r_i}{2^i}\right)\right| \geq \delta/2$$

One of the terms of this sum is greater than its average; therefore there exists a certain $\theta = \sum_{i \in S} \frac{r_i}{2^i}$ such that:

$$a_{r_1} \dots a_{r_k} \tilde{F}(\theta) \geq \frac{\delta}{2} \left(\frac{\epsilon^3}{100} \right)^k$$

Use $|a_{r_i}| < 1$ and $\epsilon = \delta/2k$:

$$\tilde{F}(\theta) \geq \frac{\delta}{2} \left(\frac{\delta^3}{200k} \right)^k > \left(\frac{\delta}{10k} \right)^{4k}$$

which is the desired bound, and proves this part. The form of θ follows naturally from this construction.

We now detail the construction of $\tilde{\psi}$ as follows. First, consider a function $\psi_1 = \phi * \chi * \chi$, where $\phi(t) = \psi(t + \epsilon/24)$, $\chi = 24/\epsilon \cdot 1_{[-\epsilon/48, \epsilon/48]}$, for some $\epsilon > 0$. The operator $*$ denotes the convolution operator, on the definition interval $[0, 1]$:

$$f * g(t) = \int_0^1 f(\tau) g(t - \tau) d\tau$$

Intuitively, this builds a continuous function that is close to ψ and can be further refined to fulfill all requirements. More specifically, the reasons for this construction are as follows:

- $\psi_1 : \mathbb{R}/\mathbb{Z} \to [-1, 1]$. This is natural, since ψ and χ are functions of period 1 with values in $[-1, 1]$.
- ψ_1 and ψ are the same almost everywhere, except on intervals whose length go to 0 as $\epsilon \to 0$. To see this, look first at $\phi * \chi$:

$$\begin{aligned}
\psi_1(t) = \phi * \chi * \chi(t) &= \int_0^1 \chi(\tau_0) \phi * \chi(t - \tau_0) d\tau_0 \\
&= \frac{24}{\epsilon} \int_{-\epsilon/48}^{\epsilon/48} \phi * \chi(t - \tau_0) d\tau_0 \\
&= \frac{24}{\epsilon} \int_{-\epsilon/48}^{\epsilon/48} \int_0^1 \chi(\tau_1) \phi(t - \tau_0 - \tau_1) d\tau_1 d\tau_0 \\
&= \left(\frac{24}{\epsilon} \right)^2 \int_{-\epsilon/48}^{\epsilon/48} \int_{-\epsilon/48}^{\epsilon/48} \psi(t - \tau_0 - \tau_1 - \epsilon/24) d\tau_1 d\tau_0
\end{aligned}$$

  Recall that the function ψ is constant on the two intervals $[0, 1/2]$ and $[1/2, 1]$. When the argument of ψ is far away enough from the points $0 = 1$ and $1/2$, the function $\psi_1 \equiv \psi$ because of the above relation. Because $-\tau_0 - \tau 1 - \epsilon/24 \in [-\epsilon/12, 0]$, this means that outside the intervals $I_1 = [\frac{1}{2} - \frac{\epsilon}{12}, \frac{1}{2}]$ and $I_2 = [1 - \frac{\epsilon}{12}, 1]$ the functions ψ and ψ_1 take the same values.

For any i, the rational numbers with denominator 2^i are evenly distributed across the interval $[0, 1]$, therefore conclude the following about the average value of the difference between ψ and ψ_0:

$$\frac{1}{N} \left| \psi_0 \left(\frac{x}{2^i} \right) - \psi \left(\frac{x}{2^i} \right) \right| \leq 2(|I_1| + |I_2|) = \frac{\epsilon}{3}$$

From the Fourier convolution theorem, the Fourier coefficients of ψ_1 are given by $\hat{\psi_1}(r) = \hat{\phi}(r) \hat{\chi}(r)^2$. Calculate the Fourier coefficients of χ directly:

$$\begin{aligned}
\hat{\chi}(r) &= \int_0^1 \chi(r) e^{2i\pi x r} dx \\
&= \frac{24}{\epsilon} \int_{-\epsilon/48}^{\epsilon/48} e^{2i\pi x r} dx \\
&= \frac{48}{\epsilon} \int_0^{\epsilon/48} \cos(2\pi x r) dx = -\frac{24}{\pi r \epsilon} \sin(\epsilon x r / 24)
\end{aligned}$$

Using the trivial bound on all Fourier coefficients $\hat{\phi}, \hat{\chi} \leq 1$ deduce the bound:

$$|\hat{\psi_1}(r)| \leq |\hat{\chi}(r)|^2 \leq \min\left(1, \frac{24}{\pi\epsilon|r|}\right)$$

and use this bound to obtain another bound on:

$$\begin{aligned}\sum_{r\geq\frac{100}{\epsilon^3}} |\hat{\psi_1}(r)| &\leq \sum_{r\geq\frac{100}{\epsilon^3}} \min\left(1, \frac{24}{\pi\epsilon|r|}\right)^2 \\ &\leq \left(\frac{24}{\pi\epsilon}\right)^2 \sum_{r\geq\frac{100}{\epsilon^3}} \frac{1}{r^2} \\ &\leq \left(\frac{24}{\pi\epsilon}\right)^2 \frac{\pi^2}{6} < \frac{\epsilon}{3}\end{aligned}$$

Of course, this bound is useful because we wish to introduce $\psi_2(r) = \sum_{r\leq 100/\epsilon^3} \hat{\psi_1}(r)\exp(2i\pi rt)$. This function is a truncation of the Fourier expansion of ψ_1 which satisfies the requirement on its Fourier series, and is also close to ψ_1 and implicitly to ψ:

$$\max_r |\psi_1(r) - \psi_2(r)| \leq \left|\sum_{r>100/\epsilon^3} \hat{\psi_1}(r)\exp(2i\pi rt)\right| \leq \epsilon/3$$

This function is almost what is required; the final function should take values in $[-1, 1]$. From the above inequality, $|\psi_2(r)| \leq \epsilon/3 + 1$, so introduce $\tilde{\psi} = \psi_2/(1+\epsilon/3)$. Again:

$$\max_r |\tilde{\psi}(r) - \psi_2(r)| \leq \max(\psi_2)\left(\frac{1}{1+\epsilon/3}\right) \leq \epsilon/3$$

and, finally, by the triangle inequality:

$$\frac{1}{N}\sum_{x<N} \left|\psi\left(\frac{x}{2^i}\right) - \tilde{\psi}\left(\frac{x}{2^i}\right)\right| \leq \epsilon$$

therefore proving that $\tilde{\psi}$ is a function satisfying the requirements.

□

By using this relation of Katai, an upper bound on the Fourier coefficients can be translated into another upper bound on the Fourier-Walsh coefficient, for some subset S. Recall that we still needed to prove 3.3:

$$\hat{\mu}(S) = O(ke^{-cn^{1/2}/k}), \text{ for } |S| = k$$

This bound is nontrivial for $k = O(n^{1/2}/\log n)$, therefore those are the interesting values that should be used with the Katai relation. Unfortunately, the existing bounds for $\hat{\mu}$ do not yield the desired result, so we have to rely on the specific sparse dyadic form of θ.

- Assuming the Generalized Riemann Hypothesis, one can deduce an upper bound $\hat{\mu}(\theta) \ll_\epsilon N^{-1/4+\epsilon}$, and this can be translated into a bound for $\hat{\mu}(S), |S| = k$.

  By way of contradiction, assume that there exists subsets S of size k such that $\hat{\mu}(S) \neq o(1)$ as $n \to \infty$, which means that $\hat{\mu}(S) > \delta$ a given constant. Therefore, by Katai's relation, there exists $\theta \in [0, 1]$ such

that:

$$\begin{aligned}\hat{\mu}(\theta) &\geq \left(\frac{\delta}{10k}\right)^{4k} \\ 2^{n(-1/4+\epsilon)} &\geq \left(\frac{\delta}{10k}\right)^{4k} \\ n(-1/4+\epsilon)\log 2 &\geq 4k\log\left(\frac{\delta}{10k}\right) \\ n &\leq ck\log k, \text{ for } c>0 \\ c'\frac{n}{\log n} &\leq k, \text{ for } c'>0\end{aligned}$$

Therefore, under this assumption, the bound holds for $k = O(n/\log n)$, which is more than enough to prove our point at hand.

- On the other hand, without assuming the GRH, the best known bound for $\hat{\mu}(\theta)$ is much higher, $|\hat{\mu}(\theta)| \ll_A \log^{-A}(N)$, for any constant $A > 0$. This bound also leads to a bound on $\hat{\mu}(S)$:

$$\begin{aligned}n^{-A}\log^{-A} 2 &\geq \left(\frac{\delta}{10k}\right)^{4k} \\ A(\log n + \log\log 2) &\leq 4k(\log 10 + \log k - \log\delta) \\ c\frac{\log n}{\log\log n} &\leq k, \text{ for } c>0\end{aligned}$$

This bound for k is insufficient, smaller than the required $k = O(n^{1/2})$.

4.2 Using the dyadic structure of θ

To make use of the structure of θ, we shall prove the following lemma:

Lemma 4.2 There exists a constant c with the following property: if $k < n^{1/2}$, and:

$$\theta = \frac{r_1}{2^{i_1}} + \cdots + \frac{r_k}{2^{i_k}} \text{ with } |r_i| < \exp(c\log^{1/2} N)$$

then:

$$|\hat{\mu}(\theta)| = O(\exp(-c\log^{1/2} N))$$

We will show that this statement proves the main point 3.3. Suppose that there exists a subset S such that $|\hat{\mu}(S)| \geq ke^{-cn^{1/2}/k}$. Then by Katai's relation, there exists a sparse dyadic $\theta\frac{r_1}{2^{i_1}} + \cdots + \frac{r_k}{2^{i_k}}$ for which $|r_i| \ll \exp(3cn^{1/2}/k)$ and $|\hat{\mu}(\theta)| \geq \exp(4cn^{1/2}/k)$. Replace $N = 2^n$ in the statement's relation and choose c small enough so that a contradiction appears.

□

Returning to the proof of **Lemma 4.2**, we shall use the following theorem:

Theorem 4.3 Let μ be the Mobius character, and consider the Dirichlet characters χ to modulus $q = 2^t \leq \exp(c_1\log^{1/2} N)$. Then:

$$\frac{1}{N}\left(\sum_{x<N}\mu(x)\chi(x)\right) = O(\exp(-c_1\log^{1/2} N))$$

We shall not prove the theorem here, since it would require a vast nonelementary setup which space restricts us from developing. As Ben Green also highlights, this bound is established by standard analytic methods such as Perron's formula and the analysis of zero-free regions for Dirichlet characters χ, and a proof can be found in [9], more specifically Exercise 11.3.7.

For our purposes, we shall use **Theorem 4.3** to prove the following corollary, which effectively settles the problem for $\theta = a/2^t$.

Corollary 4.4 For $c_2 = c_1/2$, $2^t \leq O(\exp(c_1 \log^{1/2} N))$ and $a \in \mathbb{Z}$:

$$\frac{1}{N}\left(\sum_{x<N} \mu(x) \exp(2i\pi ax/2^t)\right) = O(\exp(-c_2 \log^{1/2} N))$$

To prove this, look at the sum over all Dirichlet characters $\chi : (\mathbb{Z}/2^t\mathbb{Z})^* \to \mathbb{C}^*$ of $\chi(x)$ for a given x:

$$\sum_{\chi} \chi(x) = \begin{cases} 2^{t-1} & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

This property of Dirichlet characters χ is well known. In particular, the sum $\sum_\chi \chi(xr^{-1})$ vanishes if and only if $r = x$. Therefore, one can deduce the following equation, for $x \in (\mathbb{Z}/2^t\mathbb{Z})^*$ (i.e. odd):

$$\begin{aligned} \exp\left(2i\pi\frac{ax}{2^t}\right) &= \frac{1}{2^{t-1}} \exp\left(2i\pi\frac{ax}{2^t}\right) \sum_{\chi}\sum_{r} \chi(xr^{-1}) \\ &= \frac{1}{2^{t-1}} \sum_{r\in(\mathbb{Z}/2^t\mathbb{Z})^*} \sum_{\chi} \overline{\chi(r)} \exp\left(2i\pi\frac{ar}{2^t}\right)\chi(x) \end{aligned}$$

With this expansion of $\exp\left(2i\pi\frac{ax}{2^t}\right)$, expand the term in the sum that is to be bounded for odd x, and reverse the sums:

$$\begin{aligned} \sum_{x<N,2\nmid x} \mu(x) \exp\left(2i\pi\frac{ax}{2^t}\right) &= \sum_{x<N,2\nmid x} \mu(x) \left(\frac{1}{2^{t-1}} \sum_{r\in(\mathbb{Z}/2^t\mathbb{Z})^*} \sum_{\chi} \overline{\chi(r)} \exp\left(2i\pi\frac{ar}{2^t}\right)\chi(x)\right) \\ &= \frac{1}{2^{t-1}} \sum_{x<N,2\nmid x} \sum_{\chi} \left(\mu(x)\chi(x) \sum_{r\in(\mathbb{Z}/2^t\mathbb{Z})^*} \exp\left(2i\pi\frac{ar}{2^t}\right)\overline{\chi(r)}\right) \end{aligned}$$

Bound the absolute value of this sum, by using the trivial $\left|\exp\left(2i\pi\frac{ar}{2^t}\right)\overline{\chi(r)}\right| = 1$ for odd r and the triangle identity. Together with **Corollary 4.4** this yields the bound:

$$\begin{aligned} \sum_{x<N,2\nmid x} \mu(x) \exp\left(2i\pi\frac{ax}{2^t}\right) &\leq \frac{1}{2^{t-1}} \sum_{x<N,2\nmid x} \sum_{\chi} \left(\mu(x)\chi(x)2^{t-1}\right) \\ &\leq 2^{t-1} \sup_{\chi} \left|\sum_{x<N,2\nmid x} \mu(x)\chi(x)\right| \\ &\leq O(N \exp(-c_2 \log^{1/2} N)) \end{aligned}$$

We now deal with even values of x. This is not too hard, since $\mu(2x) = -\mu(x)$ for odd x and $\mu(2x) = 0$ for even x, therefore:

$$\begin{aligned} \sum_{x<N,2|x} \mu(x) \exp\left(2i\pi\frac{ax}{2^t}\right) &= \sum_{x<N,2|x,4\nmid x} \mu(x) \exp\left(2i\pi\frac{ax}{2^t}\right) \\ &= -\sum_{x<N/2,2\nmid x} \mu(x) \exp\left(2i\pi\frac{ax}{2^{t-1}}\right) \\ &= O(N \exp(-c_2 \log^{1/2} N)) \end{aligned}$$

Add the bounds for the two sums and arrive at the proof of the corollary.

□

The next corollary proves the problem for θ close to a number of the form $a/2^t$.

Corollary 4.5 Let $c_3 = c_2/3$ and $q = 2^t \leq \exp(c_3 \log^{1/2} N)$. If $|\theta - a/q| \leq \exp(c_3 \log^{1/2} N)/N$, then:

$$\hat{\mu}(\theta) = O(\exp(-c_3 \log^{1/2} N))$$

The main idea behind proving this corollary is approximating θ with a/q in the sum $\hat{\mu}(\theta) = \sum_{x<N} \mu(x) e^{2i\pi\theta x}$, after splitting up the interval $[1, N]$ in N/L intervals of size L, where L is to be determined. For such an interval I of length L and for $x_0 \in I$, estimate:

$$\begin{aligned}
\left|\sum_{x\in I} \mu(x) \exp(2i\pi x\theta)\right| &= \left|\exp(2i\pi x_0(\theta - a/q)) \sum_{x\in I} \mu(x) \exp\left(2i\pi \frac{ax}{q}\right) \exp(2i\pi(x - x_0)(\theta - a/q))\right| \\
&\leq \left|\sum_{x\in I} \mu(x) \exp\left(2i\pi \frac{ax}{q}\right)\right| + \left|\sum_{x\in I} \mu(x) \exp\left(2i\pi \frac{ax}{q}\right) (\exp(2i\pi(x - x_0)(\theta - a/q)) - 1)\right| \\
&\leq O(N \exp(-c_2 \log^{1/2} N)) + \sum_{x\in I} |2i\pi(x - x_0)(\theta - a/q)| \\
&\leq O(N \exp(-c_2 \log^{1/2} N)) + LO(L \exp(c_3 \log^{1/2} N)/N)
\end{aligned}$$

Therefore, the entire sum can be bounded by bounding on each of the N/L intervals:

$$\begin{aligned}
\hat{\mu}(\theta) &= \frac{1}{N}\left(\sum_{x<N} \mu(x) \exp(2i\pi x\theta)\right) \\
&= \frac{1}{N}\frac{N}{L} O(N \exp(-c_2 \log^{1/2} N)) \frac{1}{N}\frac{N}{L} LO(L \exp(c_3 \log^{1/2} N)/N) \\
&= O\left(\frac{N}{L} \exp(-c_2 \log^{1/2} N)\right) + O\left(\frac{L}{N} \exp(c_3 \log^{1/2} N)\right)
\end{aligned}$$

Choose a suitable $L = N \exp(-2c_3 \log^{1/2} N)$ such that the bound is $O\left(\exp(-c_3 \log^{1/2} N)\right)$, therefore proving the second corollary.

□

The next corollary provides a good estimation for $\hat{\mu}(\theta)$ when θ is close to a rational number with a relatively small denominator that is a power of 2. We shall state and prove a bound for $\hat{\mu}(\theta)$ when θ is not approximated by any rational number with a small denominator; this be useful later on.

Corollary 4.6 Suppose $|\hat{\mu}(\theta)| \geq \delta$. Then there exists $q \ll (\log N/\delta)^{16}$ such that:

$$|\theta - a/q| \ll (\log N/\delta)^{16} N^{-1}$$

To prove this, choose $Q = c(\log N/\delta)^{-16} N$ for $c > 0$ to be determined. Use the Dirichlet diophantine approximation theorem to conclude that there exists $a, q, 1 \leq q \leq Q$ such that:

$$\left|\theta - \frac{a}{q}\right| \leq \frac{1}{qQ} \leq \frac{1}{q^2}$$

For such a θ, theorem 13.9 from Iwaniec and Kowalski [6] proves a bound for $\hat{\mu}(\theta)$:

$$\begin{aligned}
N|\hat{\mu}(\theta)| &\ll \left(q^{1/2}N^{1/2} + q^{-1/2}N + N^{4/5}\right)^{1/2} N^{1/2} (\log N)^4 \\
\delta(\log N)^{-4} &\ll \max\left(q^{1/4}N^{-1/4}, q^{-1/4}, N^{-1/10}\right)
\end{aligned}$$

Consider each of the three cases and establish the required bound.

- If $\delta(\log N)^{-4} \ll q^{1/4}N^{-1/4}$, then, because $q \leq Q = c(\log N/\delta)^{-16}N$:

$$\begin{aligned}\delta &\ll (\log N)^4 N^{-1/4} c^{1/4} (\log N/\delta)^{-4} N^{1/4} \\ \delta^5 &\ll c^{1/4}\end{aligned}$$

  which does not hold if c is small enough.

- If $\delta(\log N)^{-4} \ll q^{-1/4}$, then

$$q \ll \frac{\log N}{\delta}$$

  as required, and the proof is over.

- If $\delta(\log N)^{-4} \ll N^{-1/10}$, then, for example, $\delta \ll N^{1/16}$ and the bound is trivially achieved, because:

$$(\log N/\delta)^{16} N^{-1} \gg (\log N)^{16} N^{-1} N \gg 1 \geq |\theta - a/q|$$

□

This result provides a bound for values of θ that are far from rationals with small denominators, and **Corollary 4.6** takes care of θ that are close to dyadic numbers. We shall prove that there are no sparse dyadic numbers that do not fall into one of these categories, and this will imply the final result.

Lemma 4.7 Let $\theta = \sum_{t=1}^{k} \frac{r_t}{2^{i_t}}$ be a sparse dyadic number, such that $i_1 < i_2 < \cdots < i_k \leq n$ and $|r_i| \leq Q$. Suppose furthermore that $|\theta - a/q| \leq Q/N$ for some $q \leq Q, (a,q) = 1$ and that $2^{n/2k} \geq 4Q^2$. Then q is a power of two.

Denote $i_0 = 0, i_{k+1} = n$, and observe that:

$$n = \sum_{t=0}^{k} i_{t+1} - i_t$$

Among these k gaps, there has to be a largest one, that is, there is a j such that $i_{j+1} - i_j \geq n/2k$. Therefore, θ is close to the sum of its first j terms; more specifically, consider $q' = 2^j$ and $a' = r_1 2^{i_j - i_1} + \cdots + r_{j-1}2^{i_j - i_{j-1}} + r_j$. The partial geometric sum can be approximated:

$$|Q - \frac{a'}{q'}| \leq Q(2^{i_{j+1}} + 2^{i_{j+2}} + \cdots + 2^{i_k}) \leq 2^{1-n/2k}\frac{Q}{q'}$$

Use that $q' \leq 2^n - n/2k$ and therefore $Q/N \leq 1/4Qq'$. The final bound:

$$\begin{aligned}\left|\frac{a}{q} - \frac{a'}{q'}\right| &\leq \left|\theta - \frac{a'}{q'}\right| + \left|\theta - \frac{a}{q}\right| \\ &\leq 2^{1-n/2k}\frac{Q}{q'} + \frac{Q}{N} \\ &\leq \frac{1}{2Qq'} + \frac{1}{4Qq'} \leq \frac{1}{Qq'} \\ &\leq \frac{1}{qq'}\end{aligned}$$

therefore $a = a', q = q'$.

□

We are now ready to complete the proof of the main statement 4.2. Recall the statement, there exists a constant c with the following property: if $k < n^{1/2}$, and:

$$\theta = \frac{r_1}{2^{i_1}} + \cdots + \frac{r_k}{2^{i_k}} \text{ with } |r_i| < \exp(c\log^{1/2} N)$$

then:

$$|\hat{\mu}(\theta)| = O(\exp(-c\log^{1/2} N))$$

Suppose $\delta = \exp(-c\log^{1/2} N)$ such that there exists θ for which $|\hat{\mu}(\theta)| \geq \delta$. According to the result we proved above, there exists $q \leq \left(\frac{\log N}{\delta}\right)^{16}$ such that the bound holds:

$$\left|\theta - \frac{a}{q}\right| \leq \left(\frac{\log N}{\delta}\right)^{16} N^{-1}$$

Because $\delta \ll 1/\log N$, we can rewrite that as:

$$\left|\theta - \frac{a}{q}\right| \leq \exp\left(\frac{-c\log^{1/2} N}{32}\right) N^{-1}$$

Now, apply the previous lemma for $Q = \frac{c\log^{1/2} N}{32}$and check that all conditions are satisfied. Indeed, $|r_i| \leq Q$ trivially by hypothesis, and $2^{n/2k} \geq 4Q^2$ as well, so q is a power of two.

Then, by the result we proved above, we get $\hat{\mu}(\theta) = O(\exp(-c_2\log^{1/2} N))$, thereby finishing the proof.

Bibliography

[1] Sarnak, P., 2011. *Three lectures on the Möbius function, randomness and dynamics. Institute for Advanced Study*, New Yersey.

[2] Tao T., 2012. *The Chowla conjecture and the Sarnak conjecture.* Unpublished. `https://terrytao.wordpress.com/2012/10/14/the-chowla-conjecture-and-the-sarnak-conjecture/`

[3] Green, B, (2012). *On (Not) Computing the Möbius Function Using Bounded Depth Circuits. Combinatorics, Probability & Computing*, 21(6), 942-951.

[4] Bender, E.A. and Goldman, J.R., 1975. On the applications of Mobius inversion in combinatorial analysis. American Mathematical Monthly, pp.789-803.

[5] Linial, N., Mansour, Y. and Nisan, N., 1993. *Constant depth circuits, Fourier transform, and learnability.* Journal of the ACM (JACM), 40(3), pp.607-620.

[6] Iwaniec, H., Kowalski, E. 2004. *Analytic Number Theory*, ISBN-10: 0-8218-3633-1, ISBN-13: 978-0-8218-3633-0, Colloquium Publications, vol. 53

[7] Hastad, J., 1986. *Almost optimal lower bounds for small depth circuits.* Proceedings of the eighteenth annual ACM symposium on Theory of computing. ACM.

[8] Razborov, A.A., 1985. *A lower bound on the monotone network complexity of the logical permanent.* Math. Zametki 37, (1985b) 887-900 (Russian) [English transl.: Math. Notes of the Acad. Sci. USSR 37(6) 485-493]

[9] Montgomery H.L., Vaughan R.C., 2006. *Multiplicative Number Theory I. Classical Theory* ISBN-13 978-0-511-25746-9

[10] Pinsky, R.G., 2014. *Problems from the Discrete to the Continuous*, Universitext, DOI 10.1007/978-3-319-07965-3_2, Springer International Publishing Switzerland

[11] Peckner, R., 2015. *Uniqueness of the measure of maximal entropy for the squarefree flow*, Israel Journal of Mathematics September 2015, Volume 210, Issue 1, pp 335-357

[12] Stuhlsatz, E., 2015 *Mobius Inversion Formula*, Unpublished, `https://www.whitman.edu/Documents/Academics/Mathematics/stuhlsatz.pdf`