UNDERGRADUATE THESIS ON BOURGAIN'S JUNTA THEOREM

WILL OVERMAN

1. Introduction

This thesis is the culmination of our work in boolean function analysis. During the course of this research project we considered a number of different questions in this area, and our main progress will be presented in section 4. In this introductory section we will give a brief overview of some of the most important definitions and tools within boolean function analysis. A much more comprehensive introduction to the area can be found in [5].

In the second section we present a result due to Bourgain [3], of which our work is an extension. We will present the entire proof as given in his original paper, but with explanations and commentary on some of the more difficult steps within the proof (at least in our opinion). In the third section we will present recent work [1] on boolean function analysis on the infinite hypercube $(\mathbb{Z}/2)^{\mathbb{N}}$, which forms a lot of the necessary machinery for our work. In the fourth section we will discuss our main result, which is that Bourgain's result can be extended to the infinite cube $(\mathbb{Z}/2)^{\mathbb{N}}$. Finally, in the last section we will discuss some of our original goals and potential future directions.

1.1. Basic definitions.

Definition 1.1. A *boolean function* is a function of the form $f : \{-1, 1\}^n \to \{-1, 1\}$

Boolean functions are abundant in computer science. Possible applications of them include the following [5]:

- May represent the desired behavior of a circuit that has n inputs and one output.
- One could identify a v-vertex graph with length $\binom{v}{2}$ strings indicating which edges are present. Then f could represent some graph property.
- In learning theory, a boolean function could represent a "concept" with n binary attributes.
- In social choice theory, a boolean function can be identified with a "voting rule" for an election with possible candidates of -1 or 1.

Definition 1.2. The *Fourier expansion* of a boolean function $f : \{-1, 1\}^n \to \{-1, 1\}$ is its representation as a real, multilinear polynomial.

The multilinear polynomial for f can have up to 2^n terms corresponding to the subsets $S \subseteq [n]$. In particular, we have $f = \sum_{S \subseteq n} \hat{f}(S) x^S$, where $x^S = \prod_{i \in S} x_i$ is the parity function of the bits in S, and where $\hat{f}(S)$ is the coefficient for x^S in the multilinear representation of f.

We can also define an inner product on pairs of functions $f, g : \{-1, 1\}^n \to \{-1, 1\}$ by $\langle f, g \rangle = 2^{-n} \sum_{x \in \{-1,1\}^n} f(x)g(x) = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x)g(x)]$.

Proposition 1.3. *For $f : \{-1,1\}^n \to \{-1,1\}$ and $S \subseteq [n]$ the Fourier coefficient of f on S is given by $\hat{f}(S) = \langle f, x^S \rangle = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x)x^S(x)]$*

Proof.

$$\begin{aligned} \langle f, x^S \rangle &= \Big\langle \sum_{T \subseteq [n]} \hat{f}(T)x^T, x^S \Big\rangle \\ &= \sum_{T \subseteq [n]} \hat{f}(T)\langle x^T, x^S \rangle \\ &= \hat{f}(S) \end{aligned}$$

□

And we can define the p-norm as $||f||_p = \mathbb{E}[|f(x)|^p]^{1/p}$.

Proposition 1.4. *Plancherel's Theorem. For any $f, g : \{-1,1\}^n \to \{-1,1\}$ we have $\langle f, g \rangle = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x)g(x)] = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S)$*

Proof.

$$\begin{aligned} \langle f, g \rangle &= \Big\langle \sum_{S \subseteq [n]} \hat{f}(S)x^S, \sum_{T \subseteq [n]} \hat{g}(T)x^T \Big\rangle \\ &= \sum_{S,T \subseteq [n]} \hat{f}(S)\hat{g}(T)\langle x^S, x^T \rangle \\ &= \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S) \end{aligned}$$

□

Proposition 1.5. *Parseval's Theorem. For $f : \{-1,1\}^n \to \{-1,1\}$, we have $\sum_{S \subseteq [n]} \hat{f}(S)^2 = 1$.*

Proof. Corollary of the previous proposition. □

The expectation of a boolean function is given by $\mathbb{E}[f] = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x) \cdot 1] = \langle f, 1 \rangle = \hat{f}(\varnothing)$. From Parseval's theorem and the expectation formula, we have that the variance is given by $\mathrm{Var}[f] = \langle f - \mathbb{E}[f], f - \mathbb{E}[f] \rangle = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{S \neq \varnothing} \hat{f}(S)^2$.

Definition 1.6. The *Fourier weight* of $f : \{-1,1\}^n \to \{-1,1\}$ on a set S is defined to be the squared Fourier coefficient $\hat{f}(S)^2$.

The Fourier weight of f at degree k is $W^k[f] = \sum_{S \subseteq [n], |S| = k} \hat{f}(S)^2$, and similary the weight above level k is given by $W^{>k}[f] = \sum_{|S|>k} \hat{f}(S)^2$.

1.2. **Influence and Hypercontractivity.** Considering $f : \{-1,1\}^n \to \{-1,1\}$ as a voting rule, we can consider the "influence" of the ith voter to be their probability of affecting the outcome. We say that a coordinate $i \in [n]$ is *pivotal* for $f : \{-1,1\}^n \to \{-1,1\}$ on input x if $f(x) \neq f(x^{\oplus i})$. Then

Definition 1.7. The *influence* of coordinate i on $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$\mathrm{Inf}_i[f] = \mathbb{P}_{x \sim \{-1,1\}^n}[f(x) \neq f(x^{\oplus i})]$$

We can combinatorially think of the influence as the fraction of dimension i edges in the Hamming cube which are boundary edges of the function, that is, edges between points with differing output values.

The influence for a coordinate can be expressed in terms of the Fourier weight as $\mathrm{Inf}_i[f] = \sum_{S \ni i} \hat{f}(S)^2$. Then the *total influence* is $I[f] = \sum_{i=1}^n \mathrm{Inf}_i[f] = \sum_{S \subseteq [n]} |S| \hat{f}(S)^2 = \sum_{k=0}^n k W^k[f]$.

2. Bourgain's theorem

One of the main motivations for this work was to study and try to use a 2002 paper by Jean Bourgain titled "On the Distribution of the Fourier Spectrum of Boolean Functions." [3] The main result in this paper, loosely speaking, is that if some boolean function is not approximately a junta of size depending on k, then we have a lower bound on the mass of the upper tail of the Fourier spectra. This section will be an overview of the paper, focused in particular on explaining relatively opaque sections of the original proof.

The main theorem of the paper is the following

Theorem 2.1. *Let $f : \{-1, 1\}^N \to \{0, 1\}$ be any boolean function. Let $k > 0$ be an integer and $\gamma > 0$ a fixed constant. Assume*

$$\sum_{|\hat{f}(S)| < \gamma 4^{-k^2}} \hat{f}(S)^2 > \gamma^2$$

then

$$W^{>k}[f] = \sum_{|S|>k} \hat{f}(S)^2 \geq c_\epsilon k^{-1/2-\epsilon}$$

for all $\epsilon > 0$, where c_ϵ is a constant depending on ϵ.

Similar to Bourgain's notation, we will write this form of an inequality as $W^{>k}[f] \gtrsim k^{-1/2}$ throughout.

In this section we will give a walk-through of this proof. We will provide discussion around some challenging aspects of the proof, which we hope will make it more approachable for some readers.

2.1. Lower bound on weight of small sets intersecting the non-"cut-off"-influential coordinates.

Begin by assuming that $\sum_{|S|>k} \hat{f}(S)^2 < \frac{\gamma^2}{100}$, which we can clearly do since otherwise we'd have a constant lower bound on the tail mass. Consider an arbitrary, fixed $\kappa \in (0, 1)$ and define the set

$$I_0 = \{i \in [N] : \sum_{S \ni i, |S| \leq k} \hat{f}(S)^2 > \kappa\}$$

Recall that $\mathbf{Inf}_i[f] = \sum_{S \ni i} \hat{f}(S)^2$. So we can see that the above set I_0 as the set of coordinates such that their "cut-off" influence, that is cutting off the contribution from sets of size above the threshold k, is larger than the parameter κ. The elements of I_0 hence depend on the choice of this parameter κ, which will be chosen in the second phase of the proof.

Since f is boolean-valued, we have Parseval's theorem $\sum_{S\subseteq[N]}\hat{f}(S)^2=1$. We can use this to get that

$$\sum_{i=1}^{N}\sum_{S\ni i,|S|\le k}|\hat{f}(S)|^2<k$$

where we also use the fact that each set will be counted at most k times since they all have to be of size at most k. Then also note that by how we defined I_0 we have

$$\kappa|I_0|\le\sum_{i=1}^{N}\sum_{S\ni i,|S|\le k}|\hat{f}(S)|^2$$

then combining these last two observations we get $|I_0|<\kappa^{-1}k$. This gives us a general bound on the size of $|I_0|$.

Now by making an assumption about the relative sizes of k and κ we can say something about the sum of the Fourier weights of sufficiently small coefficients. In particular, we will specify this in the second phase of the proof so that k and κ do satisfy this relative condition, which is as follows: If we assume $(\kappa^{-1}k)^k 16^{-k^2}<1/100$, then we have

$$\sum_{S\subseteq I_0,|S|\le k,|\hat{f}(S)|<\gamma 4^{-k^2}}\hat{f}(S)^2<(\kappa^{-1}k)^k\gamma^2 16^{-k^2} \tag{2.1}$$

$$<\gamma^2/100 \tag{2.2}$$

Let $\overline{I_0}=[N]\setminus I_0$ bet the complement of I_0. Then we consider summing over the Fourier weight of sets which intersect $\overline{I_0}$ and have size at most k. This bound will make use of three of three inequalities, the assumption in the theorem

$$\sum_{|\hat{f}(S)|<\gamma 4^{-k^2}}\hat{f}(S)^2>\gamma^2$$

our first assumption,

$$\sum_{|S|>k}|\hat{f}(S)|^2<\frac{\gamma^2}{100}$$

and the result we just established above in (2.1) and (2.2)

$$\sum_{S\subseteq I_0,|S|\le k,|\hat{f}(S)|<\gamma 4^{-k^2}}\hat{f}(S)^2<\gamma^2/100$$

Together, these all gives us

$$\sum_{S\cap\overline{I_0}\neq\varnothing,|S|\le k}\hat{f}(S)^2>\gamma^2-\gamma^2/100-\gamma^2/100 \tag{2.3}$$

$$>\gamma^2/2 \tag{2.4}$$

This is in line with the interpretation that because of the assumption of the theorem, f if not close to being a junta. Specifically, we have a set of coordinates I_- with high "cut-off" influence, and then $\overline{I_0}$ is those with less. But then this is telling us that sets which have some intersection with this set of less influential coordinates still has Fourier weight at least $\gamma^2/2$.

2.2. **Upper bound on the weight of sets intersecting a particular subset of the non-"cut-off"-influential coordinates in exactly one coordinate.**
The title of this subsection is surely very confusing for now, but hopefully will become clear throughout this subsection. By a "particular" subset of $\overline{I_0}$ we mean the following: consider a fixed subset $I_1 \subseteq \overline{I_0}$, which we will specify later. Now for a variable $x \in \{-1,1\}^N$ we write it as $x = (x_1, x_2)$ where $x_1 \in \{-1,1\}^{I_1}$. For a fixed $x_2 \in \{-1,1\}^{\overline{I_1}}$ we write $f_{x_2}(x_1)$ for $f(x_1, x_2)$. We write $F_T(x_2)$ for $\hat{f}_{x_2}(T)$. We have the Fourier expansion, where $w_T(x_1, ..., x_{|I_1|}) = (-1)^{\sum_{i \in I_1} x_i}$,

$$f(x_1, x_2) = \sum_{T \subseteq I_1} F_T(x_2) w_T(x_1)$$

Now we choose a $\delta \in (0,1)$ which we will specify later in phase two. So now we have a fixed $\kappa, \delta \in (0,1)$ and set $I_1 \subseteq \overline{I_0}$ which we have yet to specify and won't until near the second phase. Using this δ we define $\{\xi_i\}_{i \in I_1}$ to be independent $\{0,1\}$-valued "selectors" of mean $1-\delta$, by which Bourgain means that over some source of randomness ω we have $\mathbb{P}[\xi_i(\omega) = 1] = 1 - \delta$. Then we define $I(\omega) \subseteq I_1$ to be $I(\omega) = \{i \in I_1 | \xi_i(\omega) = 1\}$. Bourgain then makes the definition

$$f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}] = \sum_{T \subset I_1, T \not\subset I(\omega)} F_T(x_2) w_T(\cdot)$$

The expectation here $\mathbb{E}_{I(\omega)}[f_{x_2}]$ is in somewhat unusual notation and it's not immediately clear how it should be understood. What's going on here is that $\mathbb{E}_{I(\omega)}[f_{x_2}] : \{-1,1\}^{I_1} \to \mathbb{R}$ and is the average over uniform x' of $f_{x_2}(x_{I(\omega)}, x')$. That is, we have both a fixed x_2 and fixed values for $I(\omega)$, and then considering all possible values of the remaining bits.

Recall that f_{x_2} is $\{0,1\}$-valued. With fixed x_2 and fixed coordinates in $I(\omega)$, we have $\mathbb{E}_{I(\omega)}[f_{x_2}] = q$ for some q, implying the function takes on value 1 a fraction p of the time. This gives us the following equality

$$2 \int |f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]|^2 dx_1 = \int |f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]| dx_1 \tag{2.5}$$

since we have $2(q(1-q)^2 + (1-q)q^2) = 2q(1-q)$ on the left-hand side and $q(1-q) + (1-q)q = 2q(1-q)$ on the right-hand side.

Now fix some $1 < p < 2$, which as many of our other parameters so far, will be specified in phase two. Then we have the following

$$(p-1)^{1/2} \Big(\sum_{i \in I_1 \setminus I(\omega)} |F_{\{i\}}(x_2)|^2 \Big)^{1/2} \leq ||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_p$$

The justification for this line is actually hidden hypercontracivity. The hypercontracivity theorem tells us that for $f : \{-1,1\}^n \to \mathbb{R}$ and $1 \leq p \leq q$, we have $||T_\rho f||_q \leq ||f||_p$ for $0 \leq \rho \leq \sqrt{\frac{p-1}{q-1}}$.

In this setting, we take $q = 2$. Then $\rho = \sqrt{p-1}$. Recall $(T_\rho f)(x) = \sum_S \rho^{|S|} \hat{f}(S) x_S$. Then the above line will follow from applying the noise operator to the function obtained by just summing over the size one Fourier coefficients and applying the hypercontractivity theorem. Let $g = f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]$ and let

$$g^{=1} = \sum_{T \subset I_1, T \not\subset I(\omega), |T|=1} F_T(x_2) w_T(\cdot)$$

Then we have that

$$||T_\rho g||_2 \leq ||g||_p$$
$$||T_\rho g^{=1}||_2 \leq ||g||_p$$
$$||(p-1) \sum_{i \in I_1 \setminus I(\omega)} |F_{\{i\}}(x_2)|^2)^{1/2}||_2 \leq ||g||_p$$
$$(p-1)^{1/2}(\sum_{i \in I_1 \setminus I(\omega)} |F_{\{i\}}(x_2)|^2)^{1/2} \leq ||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_p$$

The next line to explain is, continuing with $g = f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]$,

$$||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_p \leq ||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_2^{1-2/p'} ||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_2^{2/p'}$$
$$||g||_p \leq ||g||_1^{1-2/p'} ||g||_2^{2/p'}$$

This follows from the log-convexity of L_p norms, specifically the following theorem. Let $1 \leq p_0 \leq p_1$, and let p obey $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$. Then we have $||f||_p \leq ||f||_{p_0}^{1-\theta} ||f||_{p_1}^{\theta}$. This line is just an application of this theorem with $p_0 = 1$, $p_1 = 2$, and $\theta = 2/p'$ where p' is defined by $\frac{1}{p'} = 1 - \frac{1}{p}$.

Then the next inequality to explain is

$$||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_2^{1-2/p'} ||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_2^{2/p'} \leq ||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_2^{2/p}$$
$$||g||_1^{1-2/p'} ||g||_2^{2/p'} \leq ||g||_2^{2/p}$$

This is where we use equation (2.3), $2 \int |f_{x_2} - \mathbb{E}_{J(\omega)}[f_{x_2}]|^2 dx_1 = \int |f_{x_2} - \mathbb{E}_{J(\omega)}[f_{x_2}]| dx_1$. That is, we plug in $2||g||_2^2$ for $||g||_1$. Plugging in our definition for p' and then just doing some algebra gets us to

$$\begin{aligned} ||g||_1^{1-2/p'} ||g||_2^{2/p'} &\leq (2||g||_2^2)^{1-2/p} ||g||_2^{2/p'} \\ &\leq 2^{-1+1/p} ||g||_2^{2/p} \\ &\leq ||g||_2^{2/p} \text{ since } 1 < p \\ &= ||f_{x_2} - \mathbb{E}_{J(\omega)}[f_{x_2}]||_2^{2/p} \end{aligned}$$

Then finally we can see just by the definition of $f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]$ and L_p norms

$$||f_{x_2} - \mathbb{E}_{I(\omega)}[f_{x_2}]||_2^{2/p} = [\sum_{T \subset I_1, T \not\subset I(\omega)} |F_T(x_2)|^2]^{1/p}$$

This gives us the intended result of

$$(p-1)^{1/2}(\sum_{i \in I_1 \setminus J(\omega)} |F_{\{i\}}(x_2)|^2)^{1/2} \leq [\sum_{T \subset I_1, T \not\subset I(\omega)} |F_T(x_2)|^2]^{1/p}$$

Then recalling the definition of the selectors this gives us

$$(\sum_{i \in I_1 \setminus I(\omega)} |F_{\{i\}}(x_2)|^2)^{p/2} \leq (p-1)^{-p/2} \sum_{T \subset I_1} \big(1 - \prod_{i \in T} \xi_i(\omega)\big) |F_T(x_2)|^2$$

Now we use the fact that $\int \xi_i d\omega = 1-\delta$ and Minkowski's inequality $||f+g||_p \leq ||f||_p + ||g||_p$ to get that

$$\begin{aligned}(\sum_{i\in I_1\setminus I(\omega)} |F_{\{i\}}(x_2)|^2)^{p/2} &\geq \delta^{p/2}[\sum_{i\in I_1}|F_{\{i\}}(x_2)|^2]^{p/2} - [\sum_{i\in I_1}(\xi_i(\omega)-(1-\delta))|F_{\{i\}}(x_2)|^2]^{p/2} \\ &\geq \delta^{p/2}[\sum_{i\in I_1}|F_{\{i\}}(x_2)|^2] - [\sum_{i\in I_1}(\xi_i(\omega)-(1-\delta))|F_{\{i\}}(x_2)|^2]^{p/2}\end{aligned}$$

So we have that

$$\begin{aligned}\delta^{p/2}[\sum_{i\in I_1}|F_{\{i\}}(x_2)|^2] \lesssim (p-1)^{-p/2}\sum_{T\subset I_1}\big(1-\prod_{i\in T}\xi_i(\omega)\big)|F_T(x_2)|^2 \\ +[\sum_{i\in I_1}(\xi_i(\omega)-(1-\delta))|F_{\{i\}}(x_2)|^2]^{p/2}\end{aligned}$$

We will now justify the following series of inequalities

$$\begin{aligned}&\int\int|\sum_{i\in I_1}(\xi_i(\omega)-(1-\delta))|F_{\{i\}}(x_2)|^2|^{p/2}d\omega dx_2 \\ &\leq [\int\int|\sum_{i\in I_1}(\xi_i(\omega)-(1-\delta))|F_{\{i\}}(x_2)|^2|d\omega dx_2]^{p/2} \\ &\lesssim [\int\int[\sum_{i\in I_1}(1-\xi_i(\omega))|F_{\{i\}}(x_2)|^4]^{1/2}d\omega dx_2]^{p/2}\end{aligned}$$

The first inequality follows from an application of the power means inequality, where the smaller power is $p/2 < 1$ since $1 < p < 2$ and the larger power is 1. To show the second inequality what we really need to justify is

$$\int|\sum_{i\in J}(\xi_i(\omega)-(1-\delta))|F_{\{i\}}(x_2)|^2|d\omega \lesssim \int[\sum_{i\in J}(1-\xi_i(\omega))|F_{\{i\}}(x_2)|^4]^{1/2}d\omega$$

it actually suffices to show the following, where we can get the above inequality by again using power means,

$$\int|\sum_{i\in I_1}(\xi_i(\omega)-(1-\delta))|F_{\{i\}}(x_2)|^2|d\omega \lesssim \int\sum_{i\in I_1}(1-\xi_i(\omega))|F_{\{i\}}(x_2)|^2 d\omega$$

We can understand this inequality by thinking about it in terms of a random walk, where for each point $\omega\in\Omega$ of the sample space, and for each each $i \in I_1$, we take a step of size $|F_{\{i\}}(x_2)|^2$ multiplied by a differing factor on each side of the inequality. On the left hand side, we take a step with a multiple of δ when $\xi_i(\omega) = 1$, whereas on the right hand side we simply do not take any step when $\xi_i(\omega) = 1$. And when $\xi_i(\omega) = 0$, we take a step of size $1-\delta$ to the left on the left hand size, and a step of size 1 to the right on the right hand side.

Then since we defined the ξ_i to satisfy $\int \xi_i d\omega$, we know that, for the left hand size, with probability $(1-\delta)$ we take a step of size δ to the right, and with probability δ we take a step of size $(1-\delta)$ to the left, so the expected distance of the random walk from the origin is lower than the walk on the right hand side in which with a probability δ we take a step of size 1 to the right and never take a step to the left. This establishes these equations in the paper.

Continuing on, the next thing in the proof is the following inequality,

$$
\begin{aligned}
|F_{\{i\}}(x_2)| &= |\sum_{S\cap I_1=\{i\}} \hat{f}(S)w_S(x)| \\
&\leq |\sum_{|S|\leq k, S\cap I_1=\{i\}} \hat{F}(S)w_S| + |\sum_{|S|>k, S\cap I_1=\{i\}} \hat{f}(S)w_S|
\end{aligned}
$$

and this following inequality, which makes use of the binomial theorem

$$
\begin{aligned}
[\sum_{i\in I_1} |F_{\{i\}}(x_2)|^4(1-\xi_i(\omega))]^{1/2} \leq_c &[\sum_{i\in I_1} |\sum_{|S|\leq k, S\cap I_1=\{i\}} \hat{f}(S)w_S|^4]^{1/2} \\
&+ [\sum_{i\in I_1} |\sum_{|S|>k, S\cap I_1=\{i\}} \hat{f}(S)w_S|^2](1-\xi_i(\omega))
\end{aligned}
$$

Then we integrate and make use of the Bonami lemma[2], which generally states that for $f : \{-1,1\}^n \to \mathbb{R}$ with degree k we have $||f||_4 \leq \sqrt{3}^k ||f_2||$, to get the following

$$
\begin{aligned}
&\int\int \big[\sum_{i\in I_1} F_{\{i\}}(x_2)|^4(1-\xi_i(\omega))\big]^{1/2} d\omega dx_2 \\
&\leq 3^k \big[\sum_{i\in I_1}\big(\sum_{|S|\leq k, S\cap I_1=\{i\}} |\hat{f}(S)|^2\big)^2\big]^{1/2} + \delta \sum_{|S|>k} |\hat{f}(S)|^2 \\
&\leq 3^k \max_{i\in I_1}\big(\sum_{|S|\leq k, S\ni i} |\hat{f}(S)|^2\big)^{1/2} + \delta \sum_{|S|>k} |\hat{f}(S)|^2 \\
&< 3^k \kappa^{1/2} + \delta \sum_{|S|>k} |\hat{f}(S)|^2
\end{aligned}
$$

Then combining the past few observations gives us

$$
\begin{aligned}
&\delta^{p/2} \sum_{|S\cap I_1|=1} |\hat{f}(S)|^2 \\
&\lesssim (p-1)^{-p/2} \sum_S \big[1-(1-\delta)^{|S\cap I_1|}\big] |\hat{f}(S)|^2 + \delta^{p/2}\big(\sum_{|S|>k} |\hat{f}(S)|^2\big)^{p/2} + (3^k\kappa^{1/2})^{p/2}
\end{aligned}
$$

Now we use the following bounds for the next step: if $|S\cap \overline{I_0}| \leq k$ then $1-(1-\delta)^{|S\cap I_1|} \leq \delta|S\cap I_1|$, and otherwise $1-(1-\delta)^{|S\cap I_1|} < 1$. The first inequality follows from the inequality $(1+x)^r \geq 1+rx$ for $x \geq -1$ and $r \in \mathbb{R}\setminus(0,1)$. Then we obtain the inequality

$$
\delta^{p/2} \sum_{|S\cap I_1|=1} |\hat{f}(S)|^2 \lesssim (p-1)^{-p/2}\delta \sum_S |S\cap I_1||\hat{f}(S)|^2 \tag{2.6}
$$

$$
+ (p-1)^{-p/2} \sum_{|S|>k} |\hat{f}(S)|^2 + \delta^{p/2}\big(\sum_{|S|>k} |\hat{f}(S)|^2\big)^{p/2} + (3^k\kappa^{1/2})^{p/2} \tag{2.7}
$$

This is the main inequality we will need for the desired result.

2.3. **Second phase: choosing the parameters.** We will now begin specifying a lot of the sets and quantities we had left unspecified. In particular, we will first specify $I_1 \subset \overline{I_0}$.

We fix a value t_0 in the range $0 \leq t_0 \leq \lg k$ and we let I_1 be a random subset of $\overline{I_0}$ of density $\frac{1}{1000 \cdot 2^{t_0}}$. We choose this density so that if $2^{t_0} \leq |S \cap J| < 2^{t_0+1}$, then we have

$$\mathbb{E}[|S \cap I_1|] = \frac{|S \cap \overline{I_0}|}{1000 \cdot 2^{t_0}}$$

Furthermore, for $t \geq 0$ define

$$\rho_t = \sum_{S, 2^t \leq |S \cap \overline{I_0}| < 2^{t+1}} \hat{f}(S)^2$$

then consider the expectation over the randomness we used to choose I_1 of our inequality from (2.6) and (2.7); this gives us

$$\delta^{p/2} \rho_{t_0} \lesssim (p-1)^{-p/2} \delta \big(\sum_{t \leq lgk} 2^{t-t_0} \rho_t \big) + (p-1)^{-p/2} \sum_{|S|>k} \hat{f}(S)^2$$
$$+ \delta^{p/2} \big(\sum_{|S|>k} \hat{f}(S)^2 \big)^{p/2} + (3^k \kappa^{1/2})^{p/2}$$

Now we choose some more of the parameters. Specifically take

$$\delta \sim (p-1)^{p/(2-p)} \Big(\frac{2^{t_0} \rho_{t_0}}{\sum_{t \leq \lg k} 2^t \rho_t} \Big)^{2/(2-p)}$$
$$\kappa = 10^{-k}$$

κ was chosen to make the last term on the right-hand side in the above inequality negligible, and so that our assumption of $(\kappa^{-1} k)^k 16^{-k^2} < 1/100$ from earlier in the proof is valid. Now this choice of parameters are such that our inequality implies

$$\sum_{|S|>k} \hat{f}(S)^2 \gtrsim \min \Big((p-1)^{p/(p-2)} \big(\frac{2^{t_0} \rho_{t_0}}{\sum_{t \leq \lg k} 2^t \rho_t} \big)^{2/(2-p)}, \rho_{t_0}^{2/p} \Big)$$

where $0 \leq t_0 \leq \lg k$ and $1 < p < 2$ are still arbitrary.

The rest of the proof considers two possible cases. The first we will look at is when

$$\sum_{t \leq \lg k} 2^t \rho_t < \sqrt{k}$$

Now our result from (2.1) and (2.2)

$$\sum_{S \cap \overline{I_\kappa} \neq \varnothing, |S| \leq k} \hat{f}(S)^2 > \gamma^2 - \gamma^2/100 - \gamma^2/100$$
$$> \gamma^2/2$$

is equivalent to having

$$\sum_{0 \leq t \leq \lg k} \rho_t > \gamma^2/2$$

which implies we can take a t_0, where $0 \leq t_0 \leq \lg k$, such that

$$\rho_{t_0} \gtrsim 1/\log k$$

and then we choose $p = 1 + 1/\log k$ to get that

$$\sum_{|S|>k} \hat{f}(S)^2 \gtrsim \min\left((\log k)^{-2}k^{-1/2}, (\log k)^{-2}\right) \geq_c (\log k)^{-2}k^{-1/2}$$

as desired. Then the second case is

$$\sum_{t\leq \log k} 2^t \rho_t \geq \sqrt{k}$$

then here we choose t_0 such that

$$2^{t_0}\rho_{t_0} > \frac{1}{\log k}\sum_{t\leq \log k} 2^t \rho_t > \frac{\sqrt{k}}{\log k}$$

and thus

$$\rho_{t_0} > (\log k)^{-1}k^{-1/2}$$

Now we take $p \to 2$ to get, for all $\epsilon > 0$

$$\sum_{|S|>k} \hat{f}(S)^2 \gtrsim \min\left((\log k)^{-2/(2-p)-1}k^{-1/2}, (\log k)^{-2}k^{-1/p}\right) \gtrsim k^{-1/2-\epsilon}$$

And thus we have finally reached our desired result of Theorem 2.1

Theorem. Let $f : \{-1,1\}^N \to \{0,1\}$ be any boolean function. Let $k > 0$ be an integer and $\gamma > 0$ a fixed constant. Assume

$$\sum_{|\hat{f}(S)|<\gamma 4^{-k^2}} \hat{f}(S)^2 > \gamma^2$$

then

$$W^{>k}[f] = \sum_{|S|>k} \hat{f}(S)^2 \geq c_\epsilon k^{-1/2-\epsilon}$$

for all $\epsilon > 0$, where c_ϵ is a constant depending on ϵ.

3. Boolean functions on the infinite boolean cube $(\mathbb{Z}/2)^{\mathbb{N}}$

Bourgain's original theorem, the proof of which we just saw in the last section, pertained to functions from a finite boolean cube $\{0,1\}^N$, but we would like to consider a result of this form for the infinite-dimensional Hamming cube $\{-1,1\}^\infty = \{-1,1\}^{\mathbb{N}}$.

Our approach for doing this will be to use tools developed in a 2019 master's thesis by Vilhelm Agdur [1] to show that all the lemmas or results used in Bourgain's original proof pertaining to functions from the finite cube still work for the infinite cube $(\mathbb{Z}/2)^{\mathbb{N}}$.

3.1. **Fourier analysis on the infinite cube** $(\mathbb{Z}/2)^{\mathbb{N}}$.

As discussed in Agdur's thesis, there is a general theory of Fourier analysis for functions on locally compact abelian groups. By Tychonoff's theorem we know that $\mathbb{Z}_2^\infty$ is a locally compact abelian group, and hence we have the basic results of Fourier analysis in this setting. For example, we immediately have Parseval and Planchere's theorems in this setting, which we will utilize in the last section.

3.2. **Approximating boolean functions and hypercontractivity.** We can define a noise operator on functions from the infinite cube similarly to the finite case. Specifically,

Definition 3.1. For all $p \in [1, \infty)$ and each $\rho \in [0, 1]$ we can define the *noise operator* $T_\rho : L^p(\{-1, 1\}^\infty) \to L^p(\{-1, 1\}^\infty)$ by

$$(T_\rho f)(x) = \mathbb{E}_{y \sim N_\rho(x)}[f(y)]$$

where by $y \sim N_\rho(x)$ we mean a random string y drawn as follows. For each $i \in \mathbb{N}$ independently set $y_i = x_i$ with probability ρ and to be uniformly random from $\{-1, 1\}$ with probability $1 - \rho$. The noise operator can be seen as giving a local averaging around x. We can give the Fourier expansion of $T_\rho f$ as follows:

Proposition 3.2. *The Fourier expansion of $T_\rho f$ is given by $T_\rho f = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \chi_S$*

Proof. Since T_ρ is a linear operator it suffices to consider how it acts on the parity functions

$$\begin{aligned} T_\rho \chi_S(x) &= \mathbb{E}_{y \sim N_p(x)}[y^S] \\ &= \prod_{i \in S} \mathbb{E}_{y \in N_\rho(x)}[y_i] && \text{independence of the bits} \\ &= \prod_{i \in S} (\rho x_i) && \text{since } \mathbb{E}[y_i] = \rho x_i \\ &= \rho^{|S|} \chi_S(x) \end{aligned}$$

□

Agdur was able to establish the following hypercontractivity result for functions on the infinite cube

Theorem 3.1. *For every $\rho \in [0, 1]$, we have for any function $f \in L^{1+\rho^2}(\{-1, 1\}^\infty)$ that $||T_\rho f||_2 \leq ||f||_{1+\rho^2}$.*

In order to arrive at this result the approximation operator $A_n : L^1(\{-1, 1\}^\infty) \to L^1(\{-1, 1\}^n)$ was introduced

$$(A_n f)(x) = \mathbb{E}[f | x_1, ..., x_n]$$

For $f \in L^2$ we can actually get a formula for A_n where we use the observation that A_n is the projection onto the subspace of functions that depend only on the first n bits. The characters χ_S for $S \subseteq [n]$ are an orthonormal basis for this subspace, giving us that

$$A_n f = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S$$

In order for this approximation to be useful we need for $A_n f$ to converge to f as $n \to \infty$. In order to prove that this is indeed the case we need to Doob's Martingale Convergence Theorem and Levy's Upward Theorem

Theorem 3.2. *Doob's Martingale Convergence Theorem. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, and X_k is a martingale. For all $p \in [1, \infty)$, if there exists a constant K such that $||X_k||_p \leq K$ for all k, then there exists a random variable $X \in L^p$ such thatt $X_k \to X$ almost surely in L^p.*

Theorem 3.3. *Levy's Upward Theorem. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, and $\mathcal{F}_n \nearrow \mathcal{F}$ is some filtration. Then for any $X \in L^1$, it holds that $\mathbb{E}[X|\mathcal{F}_n] \to X$ almost surely.*

Vadhan was able to use these two theorems to prove the following lemma

Lemma 3.3. *For every $p \in [1, \infty)$, it holds for every $f \in L^p(\{-1,1\}^\infty)$ that $A_n f \to f$ almost surely and in L^p.*

Proof. The filtration here will be the information gained from exposing the first n coordinates, specifically $\mathcal{F}_n$ is the sigma algebra generated by exposing the first n coordinates. We have $A_n f = \mathbb{E}[X|\mathcal{F}_n]$ so by Levy's Upward Theorem we know that $A_n f$ will converge to f almost surely.

Now we can show convergence in L^p where we first make use of Jensen's inequality

$$\begin{aligned} ||A_n f||_p^p &= \mathbb{E}[\mathbb{E}[f|\mathcal{F}_n]|^p] \\ &\leq \mathbb{E}[\mathbb{E}[|f|^p|\mathcal{F}_n]] \\ &= ||f||_p^p \end{aligned}$$

thus by just taking $K = ||f||_p$ in Doob's martingale convergence theorem, we get convergence in L^p. □

The next key lemma in extending hypercontractivity to the infinite cube is to prove that A_n and T_ρ commute as operators on L^p. Specifically, we need the following lemma

Lemma 3.4. *For all $p \in [1, \infty)$ and any $\rho \in [0,1]$, A_n and T_ρ commute as operators on L^p. So for all $F \in L^p(\{-1,1\}^\infty)$ we have*

$$A_n T_\rho f = T_\rho A_n f$$

we won't reprove this result here, but one can see from Vadhan's thesis that it takes nothing more than a argument based on conditional expectations where the conditioning is on the exposure of the first n coordinates. Now we can prove Theorem 3.1.

Proof. We begin with the finite hypercontractivity theorem, which tells us, where we treat $A_n f$ for any $n \in \mathbb{N}$ as the function, that

$$||T_\rho A_n f||_2 \leq ||A_n f||_{1+\rho^2}$$

Then by Lemma 3.4 we can use commutativity to get

$$||A_n(T_\rho f)||_2 \leq ||A_n f||_{1+\rho^2}$$

Then Lemma 3.3 implies that the right hand side of this inequality will converge, but in order to apply it to the left hand side we must argue that $T_\rho f \in L^2$. We can see this from observing that $A_n \to f$ in $L^{1+\rho^2}$ tells us that $||A_n f||_{11+\rho^2} < \infty$ and hence we can use this supremum as K in Doob's martingale convergence theorem. So then taking the limit as $n \to \infty$ of both sides gives us the desired result. □

4. Bourgain's Theorem on the infinite Boolean cube

The purpose of the previous section was to set up the tools we needed in order to carry over the argument from Bourgain's original proof to the case of functions on the infinite cube. Bpurgain's original proof essentially carries over to this new setting, but we will point out places in which we need the machinery from the previous section. The remaining parts of his original proof do not depend on the finiteness of the domain and hence we will see that his result is also true for the infinite boolean cube.

The statement of the theorem that we wish to prove for the infinite case is

Theorem 4.1. *Let $f : \{-1,1\}^\infty \to \{0,1\}$ be any boolean function. Let $k > 0$ be an integer and $\gamma > 0$ a fixed constant. Assume*

$$\sum_{|\hat{f}(S)|<\gamma 4^{-k^2}} \hat{f}(S)^2 > \gamma^2$$

then

$$W^{>k}[f] = \sum_{|S|>k} \hat{f}(S)^2 \geq c_\epsilon k^{-1/2-\epsilon}$$

for all $\epsilon > 0$, where c_ϵ is a constant depending on ϵ.

4.1. **Lower bound on weight of small sets intersecting the non-"cut-off"-influential coordinates.**
Begin by assuming that $\sum_{|S|>k} \hat{f}(S)^2 < \frac{\gamma^2}{100}$, which we can clearly do since otherwise we'd have a constant lower bound on the tail mass. Consider a, for now, arbitrary, fixed $\kappa \in (0,1)$ and define the set

$$I_0 = \{i \in \mathbb{N} : \sum_{S \ni i, |S| \leq k} \hat{f}(S)^2 > \kappa\}$$

Originally the set I_0 was just a subset of $[n]$, but we can easily define it here as a subset of the natural numbers.

As discussed in the previous section, we still have Parseval's theorem $\sum_{S \subseteq [N]} \hat{f}(S)^2 = 1$. So we can again use this to get that

$$\sum_{i=1}^{N} \sum_{S \ni i, |S| \leq k} |\hat{f}(S)|^2 < k$$

where we also use the fact that each set will be counted at most k times since they all have to be of size at most k. Then also note that by how we defined I_0 we have

$$\kappa |I_0| \leq \sum_{i=1}^{N} \sum_{S \ni i, |S| \leq k} |\hat{f}(S)|^2$$

then combining these last two observations we get $|I_0| < \kappa^{-1} k$. This gives us a general bound on the size of $|I_0|$.

The rest of the proof for this subsection is then the same as the original, since no part of the argument is affected by being in the infinite case. In particular, the remainder of this section is looking at bounds on sums over Fourier weight over certain subsets, and none of this is impacted by having a potentially infinite sum since we still have Parseval's theorem telling us that the total Fourier weight is 1.

4.2. **Upper bound on the weight of sets intersecting a particular subset of the non-"cut-off"-influential coordinates in exactly one coordinate.** Recall that in this section of the proof, the goal was to established the following, where we have some fixed $1 < p < 2$

$$(p-1)^{1/2}\Big(\sum_{i\in I_1\setminus I(\omega)}|F_{\{i\}}(x_2)|^2\Big)^{1/2}\leq ||f_{x_2}-\mathbb{E}_{I(\omega)}[f_{x_2}]||_p$$

As we said in the discussion of the original proof, the justification for this line is the hypercontracivity theorem. We know from the previous section that that hypercontracivity theorem still holds in the infinite setting, so we still have the desired inequality. Later in this section the original proof uses the Bonami lemma [2]. Since this is a corollary of the hypercontractivity theorem, it also holds here. And thus since none of the other aspects of the original proof in this section depend on the finiteness we still obtain the same main inequality

$$\delta^{p/2}\sum_{|S\cap I_1|=1}|\hat{f}(S)|^2 \lesssim (p-1)^{-p/2}\delta\sum_S |S\cap I_1||\hat{f}(S)|^2 \tag{4.1}$$

$$+(p-1)^{-p/2}\sum_{|S|>k}|\hat{f}(S)|^2+\delta^{p/2}\Big(\sum_{|S|>k}|\hat{f}(S)|^2\Big)^{p/2}+(3^k\kappa^{1/2})^{p/2} \tag{4.2}$$

4.3. **Second phase: choosing the parameters.** Similar to the original argument we simply let I_1 be a random subset of $\overline{I_0}$ of density $\frac{1}{1000\cdot 2^{t_0}}$. Since this is just a density it can be defined over $\mathbb{N}$ as it was for $[n]$. The rest of this section of the original proof does not depend on finiteness, and hence we obtain our desired result for the infinite setting

Theorem. Let $f:\{-1,1\}^\infty\to\{0,1\}$ be any boolean function. Let $k>0$ be an integer and $\gamma>0$ a fixed constant. Assume

$$\sum_{|\hat{f}(S)|<\gamma 4^{-k^2}}\hat{f}(S)^2>\gamma^2$$

then

$$W^{>k}[f]=\sum_{|S|>k}\hat{f}(S)^2\geq c_\epsilon k^{-1/2-\epsilon}$$

for all $\epsilon>0$, where c_ϵ is a constant depending on ϵ.

5. Future Directions

One of the main motivations for this work was to understand connections between Bourgain's theorem and the Levy 0-1 law. We stated this theorem earlier in section 3, but again it is

Theorem 5.1. *Levy's Upward Theorem (or the Levy 0-1 Law) Suppose* $(\Omega,\mathcal{F},\mathbb{P})$ *is some probability space, and* $\mathcal{F}_n\nearrow\mathcal{F}$ *is some filtration. Then for any* $X\in L^1$*, it holds that* $\mathbb{E}[X|\mathcal{F}_n]\to X$ *almost surely.*

Intuitively, this theorem tells us in our setting that if we have an infinite boolean function $f:\{0,1\}^\infty\to\{0,1\}$ then as we expose the coordinates of the input, we will become more and more certain as to what the value of the function on that input is. This doesn't, however, tell us anything about the *rate* of convergence to this output value. Our original goal was to use Bourgain's theorem to tell us something about this rate, which is why we had to show that it still held for infinite boolean functions $f:\{0,1\}^\infty\to\{0,1\}$.

The motivation for considering this possible connection is rooted in the fact that there is such a connection between the Kolmogorov 0-1 Law and the famous Kahn-Kalai-Linial theorem, which is the result that basically started the field of the fourier analysis of boolean functions. [4] The KKl theorem is as follows.

Theorem 5.2. *Let* $f : \{-1,1\}^n \to \{-1,1\}$*. Then there is* $i \in [n]$ *satisfying*

$$Inf_i[f] \geq \Omega(Var[f] \cdot \frac{\log n}{n})$$

Then from this theorem, we have the following lemma.

Lemma 5.1. *Let* f *be a monotone function. Then there exists* $b \in \{-1,1\}$ *and* $S \subset [n]$ *of size* $|S| \leq O(\frac{n}{\log n})$ *such that the function restricted to the subcube* $\{x : x_i = b \forall i \in S\}$*, denoted* g *satisfies* $|\mathbb{E}[g]| \geq 0.99$

Proof. Assume without loss of generality that $\mathbb{E}[f] \geq 0$ and take $b = 1$. We then have the following iterative process. Initialize $S_0 = \varnothing$. Given S_i, define g_i to be the restricted function from f in which all elements of S_i are set to 1. As long as $\mathbb{E}[g_i] < 0.99$, pick $j_i \notin S_i$ with the max influence $\text{Inf}_{j_i}[g_i]$ and set $S_{i+1} = S_i \cup \{j_i\}$. Since f is monotone we have $\mathbb{E}[g_{i+1}] = \mathbb{E}[g_i] + \text{Inf}_{j_i}[g_i] \geq \mathbb{E}[g_i] + \Omega(\frac{\lg n - i}{n-i})$, so this process must terminate after $O(\frac{n}{\lg n})$ steps since $\mathbb{E}[g_i] \leq 1$ for all i. □

And Kolmogorov's 0-1 Law tells us

Theorem 5.3. *If* X_i *is an infinite sequence of independent events and* A *is a tail event of the sequence, meaning it doesn't depend on any finite number of the events* X_i*, then* $\mathbb{P}[A] \in \{0,1\}$*.*

Then the connection between KKL and Kolmogorov can be understood as follows. think about the boolean function in the social choice setting, where it gives us some voting rule and each input x_i is an individual voter who can pick candidate 0 or 1. We are interested in swinging the election, that is, buying out a few of the voters to vote the way we want so that the output of the function is our desired candidate. Consider the contrapositive of the KKL lemma. If we can't swing the election, then the function must not depend on on this $O(\frac{1}{\lg n})$ fraction of voters, so this would be a "finite tail event".

Our goal was to find a similar sort of connection between Bourgain's result and the Levy 0-1 Law. This would be of the form of a convergence rate as we mentioned earlier. In particular, we had hoped for some quantitative form of Levy's law, possibly based on upon Bourgain's theorem on the infinite cube $(\mathbb{Z}/2)^{\mathbb{N}}$.

However, considering that majority functions Maj_n are such that $W^{>k}(\text{Maj}_n) = (2/\pi)^{3/2} k^{-1/2}(1 \pm O(1/k))$ for odd k, we can observe that since there's no dependence on n here that we cannot have a quantitative form of Levy depending only on the parameters $W^{>k}$. Regardless, finding some sort of quantitative result remains our biggest open problem.

6. Acknowledgements

I would like to thank Professor Leonard J. Schulman for his valuable mentorship, support, and guidance throughout my junior and senior years, and in particular, for advising the work on this thesis.

References

[1] V. Agdur, Finitary Boolean functions, https://arxiv.org/abs/1906.03709
[2] A. Bonami. Ensembles $\Lambda(p)$ dans le dual de D^∞. Annales de l'Institut Fourier, 18(2):193-204, 1968.
[3] J. Bourgain. On the distribution of the fourier spectrum of Boolean functions. Isr. J. Math. 131, 269-276 (2002). https://doi.org/10.1007/BF02785861
[4] J. Kahn, G. Kalai and N. Linial,The influence of variables on Boolean functions, Proc. 29th IEEE FOCS 58-80, IEEE, New York, 1988.
[5] R. O'Donnell (2014). Analysis of Boolean Functions. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139814782

California Institute of Technology, Department of Computing and Mathematical Sciences, Pasadena CA

Email address: `woverman@caltech.edu`