ON HYPERCONTRACTIVITY ON THE SYMMETRIC GROUP

UROP+ FINAL PAPER, SUMMER 2021
IBRAHIM SUAT EVREN
MENTOR: ASHWIN SAH
PROJECT SUGGESTED BY: DOR MINZER

ABSTRACT. The hypercontractive inequality is a seminal result in discrete Fourier analysis with numerous important applications in discrete mathematics. Several variants of the hypercontractivity theorem have been proved for product spaces. In a recent paper, Filmus, Kindler, Lifshitz, and Minzer proved an analagous inequality in the symmetric group setting. In this paper, we will talk about the traditional hypercontractivity theorem and its applications, and the global hypercontractivity theorem for general product spaces. Then, we shift our focus to the S_n-hypercontractive inequality and its applications. This paper is an expository paper mostly evolves around the paper by FKLM. Mainly, we develop the tools they use in their paper, and give the proof of their main theorem following one of their approaches, namely the coupling approach.

1. INTRODUCTION

In 1970, Bonami [1] proved the following theorem known as the Hypercontractivity Theorem, which is arguably the most seminal result in the field of analysis of Boolean functions.

Theorem 1.1. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be degree of d and $q \geq 2$ be a real number. Then, $\|f\|_q \leq \sqrt{q-1}^d \|f\|_2$.*

This theorem enables us to transform q-norms to 2-norm in inequalities. This is especially useful because 2-norm has a simple expression in terms of the Fourier coefficients, whereas the other norms are not as simple. In particular, $\|f\|_2^2 = \sum_{S\subseteq[n]} \widehat{f}(S)^2$. One particular simple but useful property of this identity is that each term has a nonnegative contribution. For example, we know that

$$\left\|f^{\leq d}\right\|_2^2 = \sum_{\substack{S\subseteq[n] \\ |S|\leq d}} \widehat{f}(S)^2 \leq \sum_{S\subseteq[n]} \widehat{f}(S)^2 = \|f\|_2^2$$

where $f^{\leq d}$ is defined to be the part of f whose degree is less than or equal to d.

The Hypercontractivity Theorem has been thus used to prove numerous influential results such as FKN Theorem [4] (degree-1 functions are close to a dictatorship), KKL Theorem [5] (there exists an influential voter), Invariance Principle [7] (one can switch back and forth between the Boolean setting the Gaussian setting without changing the expectation too much if the influences are small), and thereof Majority is Stablest [7] the most stable bounded function with low influences is the Majority function) and it has been used to give shorter proofs of different versions of Arrow's Impossibility Theorem (there is no "fair" voting system), including a robust version by Gil Kalai (see e.g. [8]). In fact, the theorem stated above is a corollary of a more general theorem also known as the Hypercontractivity Theorem. Before we state the original Hypercontractivity Theorem, let us give the definition of *the noise operator*, which is a linear operator from $L^2(\{-1,1\}^n)$ to itself, also known as *the smoothing operator*.

Date: September 1, 2021.
This work was funded by MIT UROP+.

Definition 1.2. Let $\rho \in [0,1]$. For any $x \in \{-1,1\}^n$ we denote by $y \sim N_\rho(x)$ the following distribution: $\forall i \in [n]$ $y_i = x_i$ with probability ρ, and uniformly random with probability $1-\rho$. The noise operator is defined to be $T_\rho f(x) = \mathbb{E}_{y\sim N_\rho(x)}[f(y)]$.

The reason T_ρ called the noise operator is that it introduces noise to each coordinate. A machine receives the signal correctly with probability ρ but randomly assigns a value to the signal by itself with probability $1-\rho$. In this sense, the functions with more variables should be affected more from this process. Indeed, $T_\rho f(x) = \sum_{S\in[n]} \rho^{|S|}\widehat{f}(S)\chi_S(x)$. So, it smooths out the high degree part of f, which is why it is called the smoothing operator. Now, we can state the Hypercontractivity Theorem in its most general version:

Theorem 1.3. *Let $f : \{-1,1\}^n \to \mathbb{R}$, $1 \le p \le q$, and $0 \le \rho \le \sqrt{\frac{p-1}{q-1}}$. Then, $\|T_\rho f\|_q \le \|f\|_p$.*

We will give the proof for $p = 2, q = 4, \rho = 1/\sqrt{3}$, which is sufficient to prove most of the applications as such. The idea behind the proof for the general case is similar, but requires more rigorous calculations. Before giving the proof, let us give the definition of discrete derivatives and Laplacians, which are two most important notions lying at the heart of many inductive proofs in discrete Fourier analysis.

Definition 1.4. Let $f : \{-1,1\}^n \to \mathbb{R}$. For every $i \in [n]$, discrete derivative with respect to coordinate i is defined to be $D_i f(x) = (f(x_{i\to 1})-f(x_{i\to -1}))/2$ where $x_{i\to 1} = (x_1,\cdots,x_{i-1},1,x_{i+1},\cdots,x_n)$ and $x_{i\to -1} = (x_1,\cdots,x_{i-1},-1,x_{i+1},\cdots,x_n)$. Laplacian is defined to be $L_i f(x) = \mathbb{E}_{x_i\sim\{-1,1\}}[f(x)]$.

Proof of Theorem 1.3. The proof follows by induction over n. Note that $T_\rho f(x) = T_\rho(D_n f(x) x_n) + T_\rho L_n f$ and that both $D_n f$ and $L_n f$ has domain $\{-1,1\}^{n-1}$. Then,

$$\begin{aligned}
\mathbb{E}[(T_\rho f)^4] &= \rho^4\mathbb{E}[(T_\rho D_n f)^4] + 6\rho^2\mathbb{E}[(T_\rho D_n f)^2(T_\rho L_n f)^2] + \mathbb{E}[(T_\rho L_n f)^4] \\
&\le \mathbb{E}[(T_\rho D_n f)^4] + 2\mathbb{E}[(T_\rho D_n f)^2(T_\rho L_n f)^2] + \mathbb{E}[(T_\rho L_n f)^4] \\
&\le \mathbb{E}[(T_\rho D_n f)^4] + 2\sqrt{\mathbb{E}[(T_\rho D_n f)^4]}\sqrt{\mathbb{E}[(T_\rho L_n f)^4]} + \mathbb{E}[(T_\rho L_n f)^4] \\
&\le \mathbb{E}[(D_n f)^2]^2 + 2\mathbb{E}[(D_n f)^2]\mathbb{E}[(L_n f)^2] + \mathbb{E}[(L_n f)^2]^2 \\
&= \mathbb{E}[f^2]^2
\end{aligned}$$

where the first equality is by binomial expansion, the second inequality is by Cauchy-Schwarz, and the third inequality is by induction. ∎

In a recent paper, Filmus, Keevash, Long, and Minzer [6] proved an analogous result for general product spaces. However, their result requires the function to be *global*.

Definition 1.5. Let $(\Omega,\mu) = (\Omega_1\times\Omega_2\times\cdots\times\Omega_m, \mu_1\times\mu_2\times\cdots\times\mu_m)$ be a product probability space and $f : \Omega \to \mathbb{R}$. We say that f is *ε-global with constant C* if for any $T \subset [m]$ and $x \in \prod_{i\in T}\Omega_i$, $\|f_{T\to x}\|_2^2 \le C^{|T|}\varepsilon$.

The statement of the Global Hypercontractivity is as follows.

Theorem 1.6. *Let $q \in \mathbb{N}$ be even, $f : \Omega \to \mathbb{R}$ be ε-global with constant C, and $\rho \le \frac{1}{(10qC)^2}$. Then, $\|T_\rho f\|_q \le \varepsilon^{\frac{q-2}{q}}\|f\|_2^{\frac{2}{q}}$.*

Following this result, Filmus, Kindler, Lifshitz, and Minzer [3] proved a similar hypercontractivity theorem in the symmetric group setting, which is probably the most common non-product group. The definition of globalness in the symmetric group setting will be given in the next chapter. For now, we encourage reader the think of it as the function is small on average even when some of its coordinates are restricted.

Theorem 1.7. *There exists $K > 0$ such that the following holds. Let $q \in \mathbb{N}$ be even, $n \geq q^{Kd^2}$. If $f : S_n \to \mathbb{R}$ is a $(2d, \varepsilon)$-global function of degree d, then $\|f\|_q \leq q^{O(d^3)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$.*

As hypercontractivity is the main tool in most of the results regarding Boolean functions, Theorem 1.7 can be used to prove analogous results in the symmetric group setting. In their paper, FKLM gives several applications of this theorem. The first application is that the symmetric group version of the well known level-d inequality (e.g. [9, Corollary 9.25]), which gives an upper bound on the low degree weights. Another one is that they prove that global product-free sets in A_n are small. In [2], Eberhard proved an upper bound without the globalness condition. FKLM's work tightens the bound. A third important result they prove is analogous to KKL Theorem [5]. They work out the proof by proving isoperimetric inequalities in the symmetric group setting for global sets. Finally, we will mention that they are able to transfer the results to other non-product domains, and the multi-slice in particular. In fact, it is not hard too see that there is a canonical correspondence between the functions of domain S_n and multi-slice. Hence the results are generally transferred in a black-box fashion.

In the rest of this paper we will develop the tools to prove Theorem 1.7. In [3], they give two different approaches. In our paper, we will stick to their first approach, which is less direct and requires the combination of a few ideas from combinatorics. Throughout the paper, we will skip some of the proofs which we think are not essential to grasp the idea and mostly straightforward.

Acknowledgements. I would like to thank Dor Minzer for suggesting the direction of the project and teaching me about the topic, Elchanan Mossel for introducing me to the field, MIT UROP+ committee for funding the project, and most importantly Ashwin Sah for being a supportive mentor throughout the project both regarding the project and else.

2. Globalness

Globalness will be a condition in the hypercontractivity theorem of S_n too. In fact, the main reason behind this requirement is that in our proof, we first transfer the function to a product space, use hypercontractivity in this product space, and transfer it back to the S_n setting. Before we give the formal definition of globalness, we first need to describe how the restrictions work in the symmetric group setting.

Given $f : S_n \to \mathbb{R}$ and a subset $T \in [n] \times [n]$ of the form $\{(i_1, j_1), (i_2, j_2), \cdots, (i_t, j_t)\}$ where all of the i's are distinct and all of the j's are distinct we denote by S_n^T the set of permutations respecting T. Similar to the Boolean setting, we denote by $f_{\to T} : S_n^T \to \mathbb{R}$ the restriction of f to S_n^T equipped with the uniform measure.

Now, we are ready to give the formal definition of globalness. We will give two different definitions and refer to the first one as full globalness, the second one as bounded globalness.

Definition 2.1. A function $f : S_n \to \mathbb{R}$ is called *ε-global with constant C* if for any consistent T, it holds that $\|f_{\to T}\|_2 \leq C^{|T|}\varepsilon$.

Definition 2.2. A function $f : S_n \to \mathbb{R}$ is called *(d, ε)-global* if for any consistent T of size at most d, it holds that $\|f_{\to T}\|_2 \leq \varepsilon$.

One can see that full globalness implies bounded globalness. So, intuitively, full globalness is stronger than bounded globalness. Indeed, we prove that bounded globalness also implies full globalness.

Lemma 2.3 ([3, Lemma 3.5])**.** *Suppose $n \geq Cd \log d$ for a sufficiently large constant C. Let $f : S_n \to \mathbb{R}$ be a $(2d, \varepsilon)$-global function of degree d. Then, f is also ε-global with constant 4^8.*

The rest of this section will be dedicated to the proof of Lemma 2.3.

2.1. **Derivatives and the Proof of Lemma 2.3.** Recall that we mentioned derivatives are crucial and lie at the heart of the proofs by induction in analysis of Boolean functions. Likewise, the notion of discrete derivatives lies at the heart of the proof of Lemma 2.3 which follows by several inductions. In the proof, we first define discrete derivatives in the symmetric group setting such that (1) if f is bounded global then the derivatives are small, (2) if derivatives are small, then f is bounded global, (3) the degree of a derivative is strictly smaller than the degree of f. Here, property (3) seems to be unrelated as we have not discussed the exact relation between small derivatives and bounded globalness in the first properties, which make it seem like only these two properties will be enough to prove the lemma. We find that the high order derivatives have exponentially greater upper bounds. Thus, the upper bound for derivatives become exponential in n without having the last property. On the other hand, we use the third property to say that high order derivatives are already 0. Therefore, the upper bound for derivatives turn out to be exponential in the degree of function rather than the number of variables in function.

Definition 2.4. Derivative of f along $(i,j) \to (k,l)$, $D_{(i,j)\to(k,l)}f : S_n^{(i,k),(j,l)} \to \mathbb{R}$ is defined as $D_{(i,j)\to(k,l)}f(\pi) = f(\pi) - f(\pi \circ (i,j))$ for every $\pi \in S_n^{(i,k),(j,l)}$. For any $t \leq n/2$, the composition of t consistent derivatives is called a derivative of order t.

Throughout the proofs of these properties of derivatives, induction will be our main tool. The following lemma shows that bounded globalness requires small derivatives. We will skip its proof as it is immediate by induction over t and the triangle inequality.

Claim 2.5 ([3, Claim 4.2])**.** *Let $t \in \mathbb{N}$, $\varepsilon > 0$, and $f : S_n \to \mathbb{R}$ be a $(2t, \varepsilon)$-global function, then for each derivative D of order t, we have $\|Df\|_2 \leq 2^t\varepsilon$.* □

Next, we will show that small derivatives imply bounded globalness.

Claim 2.6 ([3, Claim 4.2])**.** *Let $t \leq n/2, \varepsilon > 0$ and $f : S_n \to \mathbb{R}$. Assume that for all $r \leq t$ and derivative D of order r we have that $\|Df\|_2 \leq \varepsilon$. Then, f is $(t, 2^t\varepsilon)$-global.*

The proof of this lemma is again by induction over t. The inductive step is quite straightforward once the base case is known, i.e. for $t = 1$. Hence, we will prove the base case in the following claim, but leave the inductive step to the reader.

Claim 2.7. *Let $\varepsilon > 0$ and $f : S_n \to \mathbb{R}$. If $\|f\|_2 \leq \varepsilon$ and for any $i_1 \neq i_2, k_1 \neq k_2 \in [n]$, $\left\|D_{(i,j)\to(k,l)}f\right\|_2 \leq \varepsilon$, then f is $(1, 2\varepsilon)$-global.*

Proof. For any $i_1 \neq i_2, k_1 \neq k_2 \in [n]$, we get from the triangle inequality that

$$\varepsilon \geq \left\|D_{(i,j)\to(k,l)}f\right\|_2 = \|f_{i\to k,j\to l} - f_{i\to l,j\to k}\|_2 \geq \left|\|f_{i\to k,j\to l}\|_2 - \|f_{i\to l,j\to k}\|_2\right|.$$

Multiplying the inequality by $\|f_{i\to k,j\to l}\|_2 + \|f_{i\to l,j\to k}\|_2$, we get that

$$\varepsilon(\|f_{i\to k,j\to l}\|_2 + \|f_{i\to l,j\to k}\|_2) \geq \left|\|f_{i\to k,j\to l}\|_2^2 - \|f_{i\to l,j\to k}\|_2^2\right|.$$

Now, we take the average over l with the goal of eliminating one of the restrictions.

$$\varepsilon\mathbb{E}_l[\|f_{i\to k,j\to l}\|_2 + \|f_{i\to l,j\to k}\|_2] \geq \mathbb{E}_l[\left|\|f_{i\to k,j\to l}\|_2^2 - \|f_{i\to l,j\to k}\|_2^2\right|] \geq \left|\|f_{i\to k}\|_2^2 - \|f_{j\to k}\|_2^2\right|$$

where the last inequality is due to the triangle inequality. Note that by Cauchy-Schwarz,

$$\|f_{i\to k}\|_2 \geq \mathbb{E}_l[\|f_{i\to k,j\to l}\|_2^2]^{\frac{1}{2}} \geq \mathbb{E}_l[\|f_{i\to k,j\to l}\|_2]$$

and similarly $\|f_{j\to k}\|_2 \geq \mathbb{E}_l[\|f_{i\to l,j\to k}\|_2]$. Thus,

$$\varepsilon(\|f_{i\to k}\|_2 + \|f_{i\to k}\|_2) \geq \left|\|f_{i\to k}\|_2^2 - \|f_{j\to k}\|_2^2\right|.$$

Dividing the inequality by $(\|f_{i\to k}\|_2 + \|f_{i\to k}\|_2)$, we obtain that

$$\varepsilon \geq \left| \|f_{i\to k}\|_2 - \|f_{j\to k}\|_2 \right|.$$

Since $\varepsilon^2 \geq \|f\|_2^2 = \mathbb{E}_j[\|f_{j\to k}\|_2^2]$, for any k there is a j such that $\varepsilon \geq \|f_{j\to k}\|_2$. Combining with the above inequality using the triangle inequality, we get that $\|f_{i\to k}\|_2 \leq 2\varepsilon$. ☐

The proof of the following claim is straightforward.

Claim 2.8 ([3, Claim 4.3])**.** *If f is of degree d and D is a derivative of order t, then Df is of degree less than or equal to $d-t$.* ☐

With Claim 3.5, Claim 3.6, Claim 3.8 in hand, we can easily prove the following claim.

Claim 2.9 ([3, Claim 4.4])**.** *Let $f : S_n \to \mathbb{R}$ is a $(2d,\varepsilon)$-global function of degree d. Then f is $(t, 4^t\varepsilon)$-global for each $t \leq n/2$.*

Proof. By Claim 2.5, for every $t \leq d$ and derivative of degree t, $\|Df\|_2 \leq 2^t\varepsilon$. By Claim 2.8 for every $t > d$ and derivative of degree t, $\|Df\|_2 = 0$. Thus, for every $t \leq n/2$ and derivative of degree t, $\|Df\|_2 \leq 2^d\varepsilon$. The rest of the proof follows by Claim 2.6. ☐

To complete the proof of Lemma 3.3, we need to bound the norms of the restrictions with sets of size greater than $n/2$. For this, we use the naive bound that f is $(t, \|f\|_\infty)$-global for any t. As such, it is sufficient to upper bound the infinity norm.

Claim 2.10 ([3, Claim 4.5])**.** *Let f be a $(2d,\varepsilon)$-global function of degree d. Then $\|f\|_\infty \leq \sqrt{(6d)!}4^{3n}\varepsilon$.*

Proof. The proof is by induction over n. We will analyze it under two cases.

If $3d \leq n/2$, then by the previous claim we have that f is $(3d, 4^{3d}\varepsilon)$-global. Thus, for any consistent set T of size d, we have that $f_{\to T}$ is $(2d, 4^{3d})$-global. Hence,

$$\|f\|_\infty = \max_{|T|=d} \|f_{\to T}\|_\infty \leq \sqrt{(6d)!}4^{3n}\varepsilon$$

where we used the induction hypothesis for $f_{\to S}$ in the inequality.

If $3d > n/2$, then

$$\|f\|_\infty^2 = \max_\pi f(\pi)^2 \leq \sum_\pi f(\pi)^2 = n!\,\|f\|_2^2 < (6d)!\varepsilon^2.$$ ☐

Claim 2.9 and Claim 2.10 together completes the proof of Lemma 2.3. ☐

3. Coupling and the Markov Operator

In this section, we will define a *Markov operator* $T^{(\rho)}$ such that it satisfies the properties given in the following theorem.

Theorem 3.1 ([3, Theorem 1.2])**.** *For an even q and $C > 0$, there is $\rho > 0$ and an operator $T^{(\rho)} : L^2(S_n) \to L^2(S_n)$ satisfying the following two conditions:*

(1) If $f : S_n \to \mathbb{R}$ is ε-global with constant C, then $\left\|T^{(\rho)}f\right\|_q \leq \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$.

(2) There is an absolute constant K such that for all $d \leq \sqrt{\log n}/K$, it holds that the eigenvalues of $T^{(\rho)}$ corresponding to degree d functions are at least ρ^{-Kd}.

The reason we refer to it as a Markov operator is that the operator will be defined by averaging the values in the next step of a Markov chain. We construct the Markov chain by a coupling method. In the following section, we will first give a general approach, then specify it for S_n.

3.1. **General Coupling Approach.** In this section, we introduce an important technique called *coupling*. Generally speaking, coupling is a matching between two probability distributions. We want to introduce a coupling between two distributions when we have an important property in one of the distributions and we desire to transfer this property to the other. When we are to construct a coupling between two probability spaces, we aim to have a joint distribution where the marginal distributions remain same. In our case, the property we would like to transfer will be the hypercontractive property of a linear operator similar to the noise operator in the Boolean setting.

Consider two finite probability spaces X and Y, and suppose that $C(x, y)$ is a coupling between them such that the marginal distributions of x and y will corresponds to the probability distributions in X and Y, respectively. Using this coupling, we may define the averaging operators $T_{X\to Y} : L^2(X) \to L^2(Y)$ and $T_{Y\to X} : L^2(Y) \to L^2(X)$ as

$$T_{X\to Y} f(y) = \mathbb{E}_{x\sim C(\cdot,y)}[f(x)], T_{Y\to X} g(x) = \mathbb{E}_{y\sim C(x,\cdot)}[g(y)].$$

It is easy to see that by Jensen's inequality, the averaging operators are contractions with respect to L^p-norm for any $p \geq 1$. Suppose that we have a hypercontractive operator $T_Y : L^2(Y) \to L^2(Y)$. Then, if we define $T_X := T_{Y\to X}T_Y T_{X\to Y}$, T_X is also hypercontractive. For example, if $Y = \{-1, 1\}^2$ and $T_Y = T_{1/\sqrt{3}}$, then we know that $\|T_Y f\|_4 \leq \|f\|_2$ by the Hypercontractivity Theorem. Thus,

$$\|T_X f\|_4 = \|T_{Y\to X}T_Y T_{X\to Y} f\|_4 \leq \|T_Y T_{X\to Y} f\|_4 \leq \|T_{X\to Y} f\|_2 \leq \|f\|_2 .$$

3.2. **Our Coupling for S_n.** Define $L = [n]^2$ and let m be a sufficiently large number depending polynomially on n, e.g. $m = n^2$ will work. We will construct a coupling between S_n and L^m (One can take $X = S_n, Y = L^m$ in the above approach.). Our coupling is as follows:

(1) Choose $y \sim L^m$ uniformly at random.
(2) Greedily construct a set T of consistent pairs from y. That is, starting from $k = 1$ to m, we consider the k-th coordinate of y, add it to the set if it keeps the set consistent, increment k otherwise.
(3) Choose a permutation $x \in S_n^T$ uniformly random.

3.3. **Markov operator on S_n.** We are now ready to define the Markov operator $T^{(\rho)}$. Take $X = S_n$, $Y = L^m$, and define the averaging operators corresponding to the coupling we constructed above. Let T_Y be the noise operator T_ρ on the product space L^m, defined as usual: Every element is retained with probability ρ, and uniformly resampled otherwise. Note that this is equivalent to multiplying the Fourier level d with ρ^d for each d in the Boolean case. Now, we can give the hypercontractivity result of $T^{(\rho)}$ for global functions, which corresponds to the first condition of Theorem 3.1.

Theorem 3.2 ([3, Theorem 3.3])**.** *Let $q \in \mathbb{N}$ be even, $C, \varepsilon > 0$, and $\rho \leq \frac{1}{(10qC)^2}$. If $f : S_n \to \mathbb{R}$ is ε-global with constant C, then $\left\|T^{(\rho)} f\right\|_q \leq \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$.*

Proof Overview. In the proof of this theorem, we first show that if $f : S_n \to \mathbb{R}$ is ε-global with constant C, then so $g := T_{S_n\to L^m} f$. The proof of this claim is immediate by Cauchy-Schwarz and the fact that the marginal distributions in the coupling are equal to the distributions of the probability spaces S_n and L^m. Second, we note that for any X and Y coupled as in Section 4.1, $T_{X\to Y}$ is a contraction with respect to L^p-norm for any $p \geq 1$, which is immediate from Jensen's inequality. Then,

$$\left\|T^{(\rho)} f\right\|_q = \|T_{L^m\to S_n} T_\rho T_{S_n\to L^m} f\|_q \leq \|T_\rho T_{S_n\to L^m} f\|_q \leq \varepsilon^{\frac{q-2}{q}} \|T_{S_n\to L^m} f\|_2^{\frac{2}{q}} \leq \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$$

where the first and the third inequalities are due to the contractive property of the averaging operators, and the second inequality is due to Theorem 1.6. □

Another important property of the Markov operator is that we can estimate f as a polynomial of the Markov operator. Since we already know a hypercontractive inequality for this operator, namely Theorem 3.2, the next theorem will enable us to prove Theorem 1.7.

Lemma 3.3 ([3, Lemma 3.6])**.** *Let $n \geq K^{d^3} q^{-Cd}$ for a sufficiently large constant K, and let $\rho = 1/(400K^3q^2)$. Then, there exists a polynomial P satisfying $P(0) = 0$ and $\|P\| \leq q^{O(d^3)}$ such that*

$$\left\| P(T^{(\rho)})f - f \right\|_q \leq \frac{1}{\sqrt{n}} \|f\|_2$$

for every function of degree at most d.

The proof of this lemma is by spectral considerations. The proof has several steps, but each step has a mostly straightforward combinatorial proofs. Thus, we will give the overall proof and the main steps of the proof, but we will not give the detailed arguments for these steps.

First, let us provide some intuition about what are we trying to do and why. Suppose we find a polynomial as in Lemma 3.3. First, for it to happen, the eigenvalues must be sufficiently large. Otherwise, if f is an eigenfunction with small eigenvalue then $P(T^{(\rho)})f$ would be too close to 0 and not be close to f in q-norm. Second, it should have a small degree for uniformity purposes. For these reasons, we will first find a set of functions that makes *almost* an eigenbasis. From that, we will show that the eigenvalues are sufficiently large. Second, we prove that the eigenvalues are concentrated on a small number of values. And then we construct the polynomial.

3.4. **A basis for $L^2(S_n)$.** Because $L^2(S_n)$ does not have a natural basis like product spaces, we show that the most natural basis is indeed approximately a basis. For every consistent set $T \subset [n]^2$, define $v_T = \frac{1_T}{\|1_T\|_2}$ where 1_T is the indicator function of T, i.e. $1_T(\pi) = 1$ if π is consistent with T, $1_T(\pi) = 0$ otherwise. Note that $\langle v_T, v_T \rangle = 1$. The proof of the following proposition is by combinatorial considerations, which will be skipped.

Proposition 3.4. *The following properties of v_T and $T^{(\rho)}$ hold.*

1. *[3, Lemma 5.7] Let $d \leq n/2$ and $T \neq S$ be sets of size d. Then $\langle v_T, v_S \rangle \leq O(\frac{1}{n})$*
2. *[3, Proposition 5.8] There exist an absolute constant C such that or all consistent T, we have $\langle T^{(\rho)} v_T, v_T \rangle \geq (c\rho)^{|T|}$*
3. *[3, Lemma 5.9] Let $\rho \in (0, 1)$. Then for all sets $T \neq S$ of size at most $n/2$, we have $\langle T^{(\rho)} v_T, v_S \rangle = O(\frac{1}{\sqrt{n}})$.* □

By using Proposition 3.4, we can prove the following proposition.

Proposition 3.5 ([3, Proposition 5.10])**.** *Let C be a sufficiently large absolute constant. If $n \geq \rho^{-d} C^{d^2}$ and f is a d-junta, then*

$$\langle T^{(\rho)} f, f \rangle \geq \rho^{O(d)} \|f\|_2^2$$

And the following corollary is immediate.

Corollary 3.6 ([3, Corollary 5.11])**.** *Let C be a sufficiently large absolute constant. If $n \geq \rho^{-d} C^{d^2}$, then all the eigenvalues of $T^{(\rho)}$ as an operator from V_d to V_d are at least $\rho^{O(d)}$.*

3.5. **Eigenvalues are concentrated on d values.** Now, we will show that the eigenvalues of $T^{(\rho)}$ on V_d is concentrated on following d values: For any i, let $\lambda_i(\rho) = \langle T^{(\rho)} v_T, v_T \rangle$ where T is a set of size i. Note that $\lambda_i(\rho)$ does not depend on the choice of T due to symmetry.

Proposition 3.7 ([3, Lemma 5.12])**.** *Let C be a sufficiently large absolute constant. If $n \geq \rho^{-d} C^{d^2}$, then each eigenvalue of $T^{(\rho)}$ as an operator from V_d to V_d is equal to $\lambda_i(\rho) \cdot (1 \pm n^{-1/3})$.*

Proof. Let λ be an eigenvalue of $T^{(\rho)}$, and let f be a corresponding eigenfunction in $V_{[d],[d]}$. Write

$$f = \sum a_S v_S,$$

where the sum is over all $S = \{(i_1, j_1), \ldots, (i_t, j_t)\} \subseteq [d]$. Then $0 = T^{(\rho)} f - \lambda f$, but on the other hand for each set S we have

$$\begin{aligned}\langle T^{(\rho)} f - \lambda f, v_S\rangle &= a_S \left(\langle T^{(\rho)} v_S, v_S\rangle - \lambda\right) \pm \sum_{|S| \neq |T|} |a_T| \left(\left|\langle T^{(\rho)} v_T, v_S\rangle\right| + |\lambda| \left|\langle v_T, v_S\rangle\right|\right) \\ &= a_S \left(\lambda_{|S|}(\rho) - \lambda\right) \pm O\left(\frac{\sum_{T \neq S} |a_T|}{\sqrt{n}}\right).\end{aligned}$$

Thus, for all S we have that

$$|a_S| \left|\lambda_{|S|}(\rho) - \lambda\right| \leq O\left(\frac{\sum_{T \neq S} |a_T|}{\sqrt{n}}\right).$$

On the other hand, choosing S that maximizes $|a_S|$, we find that $|a_S| \geq \frac{\sum_{T \neq S} |a_T|}{2^{d^2}}$, and plugging that into the previous inequality yields that $\left|\lambda_{|S|}(\rho) - \lambda\right| \leq \frac{O\left(2^{d^2}\right)}{\sqrt{n}} \leq n^{-0.4} \rho^{-d} \leq n^{-1/3} \lambda_{|S|}(\rho)$, provided that C is sufficiently large. □

3.6. **Finishing the proof of Lemma 3.3.** We first prove the L^2 variant of the lemma, and then we prove the lemma itself.

Claim 3.8 ([3, Lemma 5.13])**.** *Let $n \geq \rho^{-Cd^3}$ for a sufficiently large constant C. There exists a polynomial $P(z) = \sum_{i=1}^{k} a_i z^i$, such that $\|P\| \leq \rho^{-O(d^3)}$ and $\|P\left(\mathrm{T}^{(\rho)}\right) f - f\|_2 \leq n^{-2d} \|f\|_2$.*

Proof. Choose $P(z) = 1 - \prod_{i=1}^{d} \left(\lambda_i^{-1} z - 1\right)^{9d}$, where $\lambda_i = \lambda_i(\rho)$. Orthogonally decompose $\mathrm{T}^{(\rho)}$ to write $f = \sum_\lambda f^{=\lambda}$, for nonzero orthogonal functions $f^{=\lambda} \in V_d$ satisfying $\mathrm{T}^{(\rho)} f^{=\lambda} = \lambda f^{=\lambda}$, and let $g = P\left(\mathrm{T}^{(\rho)}\right) f - f$. Then $g = \sum_\lambda \left(P(\lambda) - 1\right) f^{=\lambda}$. Therefore

$$\|g\|_2^2 = \sum_\lambda \left(P(\lambda) - 1\right)^2 \|f^{=\lambda}\|_2^2 \leq \max_\lambda (P(\lambda) - 1)^2 \|f\|_2^2.$$

Suppose the maximum is attained at $\lambda_\star$. By Proposition 3.7, there is $i \leq d$ such that $\lambda_\star = \lambda_i(1 \pm n^{-\frac{1}{3}})$, and so

$$\left|\left(\lambda_i^{-1} \lambda_\star - 1\right)^{9d}\right| \leq n^{-3d}.$$

For any $j \neq i$, we have by Corollary 3.6 that $\lambda_j \geq \rho^{O(d)}$, and so

$$\left|\left(\lambda_i^{-1} \lambda_\star - 1\right)^{9d}\right| \leq \rho^{-O(d^2)}.$$

Combining the two inequalities, we get that

$$(1 - P(\lambda_\star))^2 \leq \rho^{-O(d^3)} n^{-6d} \leq n^{-2d},$$

where the last inequality follows from the lower bound on n. Now all we need to do is to upper bound $\|P\|$ to end the proof.

$$\|P\| \leq 1 + \left\|\prod_{i=1}^{d} \left(\lambda_i^{-1} z - 1\right)^{9d}\right\| \leq 1 + \prod_{i=1}^{d} \left\|\lambda_i^{-1} z - 1\right\|^{9d} = 1 + \prod_{i=1}^{d} (1 + \lambda^{-1})^{9d} \leq 1 + \prod_{i=1}^{d} (1 + \rho^{-O(d)})^{9d},$$

which is at most $\rho^{-O(d^3)}$. In the second inequality, we used the fact that $\|P_1 P_2\| \leq \|P_1\| \|P_2\|$. □

Now, we will prove a hypercontractive inequality (which is weaker than the hypercontractive inequality we are trying to prove), and obtain the Lemma 3.3 together with the previous claim.

Lemma 3.9 ([3, Lemma 5.14])**.** *Let C be a sufficiently large absolute constant, and let $n \geq C^{d^2} q^{2d}$. Let $f : S_n \to \mathbb{R}$ be a function of degree d. Then, $\|f\|_q \leq q^{O(d)} n^d \|f\|_2$.*

Proof. Let $\rho = \frac{1}{(10 \cdot 4^8 \cdot q)^2}$. Decomposing f into the $\sum_\lambda f_{=\lambda}$ where $T^{(\rho)} f_{=\lambda} = \lambda f_{=\lambda}$, we may find g of degree d, such that $f = \mathrm{T}^{(\rho)} g$, namely, $g = \sum_\lambda \lambda^{-1} f_{=\lambda}$. By Parseval and Corollary 3.6, we get that $\|g\|_2 \leq \rho^{-O(d)} \|f\|_2$. Thus, we have that $\|f\|_q = \left\|T^{(\rho)} g\right\|_q$, and to upper bound this norm we intend to use Theorem 3.2, and for that we need to show that g is global with fairly weak parameters.

Let T be a consistent set of size $2d$. Then,

$$\|g_{\to T}\|_2^2 = \frac{\mathbb{E}_{x\sim S_n}[g(x)^2 1_T(x)]}{\mathbb{E}_{x\sim S_n}[1_T(x)]} \leq \frac{\mathbb{E}_{x\sim S_n}[g(x)^2 1_T(x)^2]^{1/2}}{\mathbb{E}_x[1_T(x)]} \mathbb{E}_{x\sim S_n}[g(x)]^{1/2} = \frac{\|g_{\to T}\|_2}{E_{x\sim S_n}[1_T(x)]^{1/2}} \|g\|_2 .$$

Thus,

$$\|g_{\to T}\|_2 \leq \frac{\|g\|_2}{E_{x\sim S_n}[1_T(x)]^{1/2}} \leq n^d \|g\|_2 \leq n^d \rho^{-O(d)} \|f\|_2 .$$

Hence, g is $(2d, n^d \rho^{-O(d)} \|f\|_2)$ global. Lemma 2.3 implies that g is $n^d \rho^{-O(d)} \|f\|_2$ global with constant 4^8. By the choice of ρ, we may use Theorem 3.2 to get that

$$\|f\|_q = \left\|T^{(\rho)} g\right\|_q \leq (n^d \rho^{-O(d)} \|f\|_2)^{\frac{q-2}{q}} \|g\|_2^{\frac{2}{q}} \leq n^d \rho^{-O(d)} \|f\|_2 = n^d q^{O(d)} \|f\|_2$$

which completes the proof. ☐

Now, we are ready to finish the proof of Lemma 3.3.

Proof of Lemma 3.3. Let f be a function of degree d. By lemma 3.8 there exists a P with $\|P\| \leq \rho^{-O(d^3)}$ and $P(0) = 0$ such that the function $g = P\left(\mathrm{T}^{(\rho)}\right) f - f$ satisfies $\|g\|_2 \leq n^{-2d} \|f\|_2$. By Lemma 3.9, $\|g\|_q \leq q^{4d} n^{-d} \|f\|_2 \leq \frac{1}{\sqrt{n}} \|f\|_2$, provided that C is sufficiently large, completing the proof. ☐

4. Proof of Theorem 1.7

We recall Theorem 1.7.

Theorem 1.7. *There exists $K > 0$ such that the following holds. Let $q \in \mathbb{N}$ be even, $n \geq q^{Kd^2}$. If $f : S_n \to \mathbb{R}$ is a $(2d, \varepsilon)$-global function of degree d, then $\|f\|_q \leq q^{O(d^3)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$.*

Proof. Let $\rho = 1/(400K^3 q^2)$, P and K be as in Lemma 3.3. Then, by the triangle inequality and Lemma 3.3,

$$\|f\|_q \leq \left\|P(T^{(\rho)}) f\right\|_q + \frac{1}{\sqrt{n}} \|f\|_2 .$$

Also,

$$\left\|P(T^{(\rho)}) f\right\|_q = \left\|\sum (T^{(\rho)})^i f\right\|_q \leq |a_i| \left\|\sum (T^{(\rho)})^i f\right\|_q \leq \|P\| \left\|(T^{(\rho)}) f\right\|_q \leq q^{O(d^3)} \left\|(T^{(\rho)}) f\right\|_q$$

where the first inequality is due to the triangle inequality, and the second inequality is because $T^{(\rho)}$ is a contraction. Given K is sufficiently large, by Lemma 2.3 and Theorem 3.2, we get that $\left\|T^{(\rho)} f\right\|_q \leq \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$. Because $\|f\|_2 \leq \varepsilon$, we have

$$\|f\|_q \leq q^{O(d^3)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}} + \frac{\varepsilon}{\sqrt{n}} = q^{O(d^3)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}} ,$$

which completes the proof. ∎

References

[1] Aline Bonami, *Étude des coefficients de fourier des fonctions de $L^p(G)$*, Annales de l'Institut Fourier. **20** (1970), 335–402.

[2] Sean Eberhard, *Product mixing in the alternating group*, Discrete Anal. (2016), Paper No. 2, 19.

[3] Yuval Filmus, Guy Kindler, Noam Lifshitz, and Dor Minzer, *Hypercontractivity on the symmetric group*, 2020, arXiv:2009.05503.

[4] Ehud Friedgut, Gil Kalai, and Assaf Naor, *Boolean functions whose Fourier transform is concentrated on the first two levels*, Advances in Applied Mathematics **29** (2002), 427–437.

[5] J. Kahn, G. Kalai, and N. Linial, *The influence of variables on Boolean functions*, 29th Annual Symposium on Foundations of Computer Science (1988), 68–80.

[6] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer, *Forbidden intersections for codes*, 2021, arXiv:2103.05050.

[7] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences: Invariance and optimality*, Annals of mathematics. **171** (2010-3-17), 295–341.

[8] Noam Nisan, *From Arrow to Fourier*, Mar 2009, https://agtb.wordpress.com/2009/03/31/from-arrow-to-fourier/.

[9] Ryan O'Donnel, *Analysis of Boolean functions*, Cambridge University Press, 2014.

Department of Mathematics, Massachusetts Institute of Technology
Email address: `evrenis@mit.edu`