Hypercontractivity on the symmetric group

Yuval Filmus [*] Guy Kindler [†] Noam Lifshitz [‡] Dor Minzer [§]

Abstract

The hypercontractive inequality is a fundamental result in analysis, with many applications throughout discrete mathematics, theoretical computer science, combinatorics and more. So far, variants of this inequality have been proved mainly for product spaces, which raises the question of whether analogous results hold over non-product domains.

We consider the symmetric group, S_n, one of the most basic non-product domains, and establish hypercontractive inequalities on it. Our inequalities are most effective for the class of *global functions* on S_n, which are functions whose 2-norm remains small when restricting $O(1)$ coordinates of the input, and assert that low-degree, global functions have small q-norms, for $q > 2$.

As applications, we show:

1. An analog of the level-d inequality on the hypercube, asserting that the mass of a global function on low-degrees is very small. We also show how to use this inequality to bound the size of global, product-free sets in the alternating group A_n.
2. Isoperimetric inequalities on the transposition Cayley graph of S_n for global functions, that are analogous to the KKL theorem and to the small-set expansion property in the Boolean hypercube.
3. Hypercontractive inequalities on the multi-slice, and stability versions of the Kruskal–Katona Theorem in some regimes of parameters.

1 Introduction

The hypercontractive inequality is a fundamental result in analysis that allows one to compare various norms of low-degree functions over a given domain. A notorious example is the Boolean hypercube $\{0,1\}^n$ equipped with the uniform measure, in which case the inequality states that for any function $f\colon \{0,1\}^n \to \mathbb{R}$ of degree at most d, one has that $\|f\|_q \leqslant \sqrt{q-1}^d \|f\|_2$ for any $q \geqslant 2$. (Here and throughout the paper, we use expectation norms, $\|f\|_q = \mathbb{E}_x\left[|f(x)|^q\right]^{1/q}$, where the input distribution is clear from context, uniform in this case). While the inequality may appear technical and mysterious at first sight, it has proven itself as remarkably useful, and lies at the heart of numerous important results, e.g. [15, 11, 2, 23].

While the hypercontractive inequality holds for general product spaces, in some important cases it is very weak quantitatively. Such cases include the p-biased cube for $p = o(1)$, the multi-cube $[m]^n$ for $m = \omega(1)$, and the bilinear graph (closely related to the Grassmann graph). This quantitative deficiency causes various analytical and combinatorial problems on these domains to be considerably more challenging, and indeed much less is known there (and what is known is considerably more difficult to prove, see for example [12]).

[*]Department of Computer Science, Technion, Israel. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 802020-ERC-HARMONIC.

[†]Einstein Institute of Mathematics, Hebrew University of Jerusalem.

[‡]Einstein Institute of Mathematics, Hebrew University of Jerusalem.

[§]Department of Mathematics, Massachusetts Institute of Technology.

1.1 Global hypercontractivity

Recently, initially motivated by the study of PCPs (probabilistically checkable proofs) and later by sharp-threshold results, variants of the hypercontractive inequality have been established in such domains [20, 17, 18]. In these variants, one states an inequality that holds for all functions, but is only meaningful for a special (important) class of functions, called *global functions*. Informally, a function f on a given product domain $\Omega = \Omega_1 \times \cdots \times \Omega_n$ is called global, if its 2-norm, as well as the 2-norms of all its restrictions, are all small.[1] This makes these variants applicable in cases that were previously out of reach, leading to new results, but at the same time harder to apply, since one has to make sure it is applied to a global function to get a meaningful bound (see [17, 22, 18] for example applications.). It is worth noting that these variants are in fact generalizations of the standard hypercontractive inequality, since one can easily show that in domains such as the Boolean hypercube, all low-degree functions are global.

By now, there are various proofs of the above mentioned results: (1) a proof by reduction to the Boolean hypercube, (2) a direct proof by expanding $\|f\|_q^q$ (for even q's), (3) an inductive proof on n.[2] All of these proofs use the product structure of the domain very strongly, and therefore it is unclear how to adapt them beyond the realm of product spaces.

1.2 Hypercontractivity on non-product spaces

Significant challenges arise when trying to analyze non-product spaces. The simplest examples of such spaces are the slice and multi-slice, and the symmetric group. The classical hypercontractive inequality is equivalent to another inequality, the log-Sobolev inequality. Sharp Log-Sobolev inequalities were proven for the symmetric group and the slice by Lee and Yau [21], and for the multi-slice by Salez [26] (improving on earlier work of Filmus, O'Donnell and Wu [10]).

While such log-Sobolev inequalities are useful for balanced slices and multi-slices, their usefulness for domains such as the symmetric group is limited, due to the similarity between S_n and $[n]^n$. For this reason, Diaconis and Shahshahnai [4] resorted to representation-theoretic techniques in their analysis of the convergence of Markov chains on S_n. We rectify this issue in a different way, by extending global hypercontractivity to S_n.

1.3 Main results

The main goal of this paper is to study the symmetric group S_n, which is probably the most fundamental non-product domain. Throughout this paper, we will consider S_n as a probability space equipped with the uniform measure, and use expectation norms, as well as the corresponding expectation inner product, according to the uniform measure. We will think of S_n as a subset of $[n]^n$, and thereby for $\pi \in S_n$ refer to $\pi(1)$ as "the first coordinate of the input".

To state our main results, we begin with defining the notion of globalness on S_n. Given $f\colon S_n \to \mathbb{R}$ and a subset $L \subseteq [n] \times [n]$ of the form $\{(i_1, j_1), \ldots, (i_t, j_t)\}$, 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 $\pi \in S_n$ respecting T (i.e. such that $\pi(i_\ell) = j_\ell$ for all

[1] We remark that this requirement can often be greatly relaxed: (1) it is often enough to only consider restrictions that fix $O(1)$ of the coordinates of the input, and (2) it is often enough that there are "very few" restrictions that have large 2-norm, for an appropriate notion of "very few".

[2] This inductive proof is actually much trickier than the textbook proof of the hypercontractive inequality over the Boolean cube. The reason is that the statement of the result itself does not tensorize, thus one has to come up with an alternative, slightly stronger, statement, that does tensorize

$\ell = 1, \ldots, t$), sometimes known as a *double coset* (and corresponding to the notion of link in complexes). We denote by $f_{\to T} \colon S_n^T \to \mathbb{R}$ the restriction of f to S_n^T, and equip S_n^T with the uniform measure.

Definition 1.1. *A function $f \colon S_n \to \mathbb{R}$ is called ε-*global with constant C *if for any consistent T, it holds that $\|f_{\to T}\|_2 \leqslant C^{|T|}\varepsilon$.*

Our basic hypercontractive inequality is concerned with a Markov operator $\mathrm{T}^{(\rho)}$ that may at first not seem very natural. We defer the precise development and motivation for $\mathrm{T}^{(\rho)}$ to Section 1.5; for now, we encourage the reader to think of it as averaging after a long random walk on the transpositions graph, say of length $\Theta(n)$.[3]

Theorem 1.2. *For an even $q \in \mathbb{N}$ and $C > 0$, there is $\rho > 0$ and an operator $\mathrm{T}^{(\rho)} \colon L^2(S_n) \to L^2(S_n)$ satisfying:*

1. *If $f \colon \{0,1\}^n \to \mathbb{R}$ is ε-global with constant C, then $\left\|\mathrm{T}^{(\rho)} f\right\|_q \leqslant \varepsilon^{\frac{q-2}{2}} \|f\|_2^{2/q}$.*
2. *There is an absolute constant K, such for all $d \in \mathbb{N}$ satisfying $d \leqslant \sqrt{\log n}/K$, it holds that the eigenvalues of $T^{(\rho)}$ corresponding to degree d functions are at least $\rho^{-K\cdot d}$.*

As is often the case, once one has a hypercontractive inequality involving a noise operator whose eigenvalues are well-understood, one can state a hypercontractive inequality for low-degree functions. For us, however, it will be important to relax the notion of globalness appropriately, and we therefore consider the notion of bounded globalness.

Definition 1.3. *A function $f \colon S_n \to \mathbb{R}$ is called (d,ε)-global if for any consistent T of size at most d, it holds that $\|f_{\to T}\|_2 \leqslant \varepsilon$.*

A natural example of (d,ε)-global functions is the *low-degree part* of f, denoted by $f^{\leqslant d}$, which is the degree d function which is closest to f in L_2-norm. Here, a function has degree d if it can be written as a linear combination of indicators of sets S_n^T for $|T| \leqslant d$. Naively, one may expect such connection to trivially hold (by Parseval); the issue is that restrictions and degree-truncations do not commute as well as in product spaces, so such naive arguments fail. Nevertheless, we show that such a connection indeed holds.

With Definition 1.3 in hand, we can now state our hypercontractive inequality for low-degree functions.

Theorem 1.4. *There exists $K > 0$ such that the following holds. Let $q \in \mathbb{N}$ be even, $n \geqslant q^{K\cdot d^2}$. If f is a $(2d,\varepsilon)$-global function of degree d, then $\|f\|_q \leqslant q^{O\left(d^3\right)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$.*

Remark 1.5. *The focus of the current paper is on the case that n is very large in comparison to the degree d, and therefore the technical conditions imposed on n in Theorems 1.2 and 1.4 will hold for us. It would be interesting to relax or even remove these conditions altogether, and we leave further investigation to future works.*

1.4 Applications

We present some applications of Theorem 1.2 and Theorem 1.4, as outlined below.

[3] Formally, our applications only require that the eigenvalues corresponding to low-degree functions are bounded away from 0 (given that n is large enough in comparison the degree of f), which will be the case.

1.4.1 The level-d inequality

Our first application is concerned with the weight a global function has on its low degrees, which is an analog of the classical level-d inequality on the Boolean hypercube (e.g. [24, Corollary 9.25]).

Theorem 1.6. *There exists an absolute constant $C > 0$ such that the following holds. Let $d, n \in \mathbb{N}$ and $\varepsilon > 0$ such that $n \geqslant 2^{Cd^3} \log(1/\varepsilon)^{Cd}$. If $f\colon S_n \to \mathbb{Z}$ is $(2d, \varepsilon)$-global, then $\left\|f^{\leqslant d}\right\|_2^2 \leqslant 2^{C\cdot d^4} \varepsilon^4 \log^{C\cdot d}(1/\varepsilon)$.*

Theorem 1.6 should be compared to the level-d inequality on the hypercube, which asserts that for any function $f\colon \{0,1\}^n \to \{0,1\}$ with $\mathbb{E}[f] = \delta < 1/2$ we have that $\left\|f^{\leqslant d}\right\|_2^2 \leqslant \delta^2 \left(\frac{10\log(1/\delta)}{d}\right)^d$, for all $d \leqslant \log(1/\delta)$. (Quantitatively, the parameter δ should be compared to ε^2 in Theorem 1.6 due to normalization).

Note that it may be the case that ε in Theorem 1.6 is much larger than $\|f\|_2^{1/2}$, and then Theorem 1.6 becomes trivial.[4] Fortunately, we can prove a stronger version of Theorem 1.6 for functions f whose 2-norm is not exponentially small, which actually follows relatively easily from Theorem 1.6.

Theorem 1.7. *There exists an absolute constant $C > 0$ such that the following holds. Let $d, n \in \mathbb{N}$, $\varepsilon > 0$ be parameters and let $f\colon S_n \to \mathbb{Z}$ be a $(2d, \varepsilon)$-global function. If $n \geqslant 2^{Cd^3} \log(1/\|f\|_2)^{C\cdot d}$, then*

$$\left\|f^{\leqslant d}\right\|_2^2 \leqslant 2^{C\cdot d^4} \|f\|_2^2 \varepsilon^2 \log^{C\cdot d}(1/\|f\|_2^2).$$

On the proof of the level-d inequality. In contrast to the case of the hypercube, Theorem 1.6 does not immediately follow from Theorem 1.2 or Theorem 1.4, and requires more work, as we explain below. Recall that one proof of the level-d inequality on the hypercube proceeds, using hypercontractivity, as

$$\left\|f^{\leqslant d}\right\|_2^2 = \langle f^{\leqslant d}, f\rangle \leqslant \left\|f^{\leqslant d}\right\|_q \|f\|_{1+1/(q-1)} \leqslant \sqrt{q-1}^{\,d} \left\|f^{\leqslant d}\right\|_2 \|f\|_{1+1/(q-1)},$$

choosing suitable q, and rearranging. Our hypercontractive inequality does not allow us to make the final transition, and instead only tells us that $\left\|f^{\leqslant d}\right\|_q \leqslant O_{d,q}(\varepsilon^{(q-2)/q}) \left\|f^{\leqslant d}\right\|_2^{2/q}$. Executing this plan only implies, at best, the quantitatively weaker statement that $\left\|f^{\leqslant d}\right\|_2^2 \leqslant \varepsilon^{3/2} \log^{O_d(1)}(1/\varepsilon)$. Here, the difference between $\varepsilon^{3/2}$ and ε^2 is often crucial, because such results are often only useful for very small ε anyway.

To explain how we circumvent this issue, note first that the source of the inefficiency is that we used the fact that $f^{\leqslant d}$ is $(2d, \varepsilon)$-global, but the reality could be that it is much more global than that (for example, the statement itself asserts a much stronger bound on the 2-norm of $f^{\leqslant d}$). To exploit this point, let us consider the restriction that maximizes the 2-norm of $f^{\leqslant d}$. The most optimistic case would be that the globalness of $f^{\leqslant d}$ is achieved already by the function itself, which would say that $f^{\leqslant d}$ is $(2d, O_d(\left\|f^{\leqslant d}\right\|_2))$-global. In this case, the argument from the hypercube goes through well enough to achieve the desired bound.

What if the globalness of $f^{\leqslant d}$ is achieved by a restriction of size r instead? In this case, we show that there is a "derivative"[5] g of $f^{\leqslant d}$ which achieves roughly the same 2-norm as that restriction of $f^{\leqslant d}$, and taking any further "derivatives" only decreases the 2-norm of g. We show that this implies that g is $(2d, O_d(\|g\|_2))$ global, so we have reached the same situation as before!

The above discussion motivates an inductive approach, and in particular proving the statement for all integer-valued functions (and not only Boolean functions), as stated. This way, we are able to show that

[4] Parseval's identity implies that the sum of all $\|f^{=d}\|^2$ is $\|f\|_2^2$, so in particular $\left\|f^{\leqslant d}\right\|_2^2 \leqslant \|f\|_2^2$.

[5] We only define the appropriate notion of derivatives we use in Section 4, and for now encourage the reader to think of it as an analogous operation to the discrete derivative in the Boolean hypercube.

for g above we have that $\|g\|_2 = \tilde{O}_d(\varepsilon^2)$, which implies that $f^{\leqslant d}$ is $(2d, \tilde{O}_d(\varepsilon^2))$-global. This is a major improvement over our original knowledge regarding $f^{\leqslant d}$, and in particular it allows us to run the argument from the hypercube (described above) successfully.

1.4.2 Global product-free sets are small

We say that a family of permutations $F \subseteq S_n$ is product-free if there are no $\pi_1, \pi_2, \pi_3 \in S_n$ such that $\pi_3 = \pi_2 \circ \pi_1$. What is the size of the largest product-free family F?

With the formulation above, one can of course take F to be the set of odd permutations, which has size $|S_n|/2$. What happens if we forbid such permutations, i.e. only consider families of even permutations?

Questions of this sort generalize the well-studied problem of finding arithmetic sequences in dense sets. More relevant to us is the work of Gowers [14], which studies this problem for a wide range of groups (referred therein as "quasi-random groups"), and the work of Eberhard [6] which specialized this question to A_n, and improves Gowers' results. More specifically, Gowers' result shows that a product-free $F \subseteq A_n$ has size at most $O\left(\frac{1}{n^{1/3}}|A_n|\right)$, and Eberhard's work [6] improves this bound to $|F| = O\left(\frac{\log^{7/2} n}{\sqrt{n}}|A_n|\right)$. We remark that Eberhard's result is tight up to the polylogarithmic factor, as can be evidenced from the family

$$F = \left\{\pi \in A_n \mid \pi(1) \in \left\{2, \ldots, \sqrt{n}\right\}, \pi(\left\{2, \ldots, \sqrt{n}\right\}) \subseteq [n] \setminus [\sqrt{n}]\right\}. \tag{1}$$

In this section, we consider the problem of determining the maximal size of a *global*, product-free set in A_n. In particular, we show:

Theorem 1.8. *There exists $N \in \mathbb{N}$ such that the following holds for all $n \geqslant N$. For every $C > 0$ there is $K > 0$, such that if $F \subseteq A_n$ is product-free and is $(6, C\cdot\sqrt{\delta})$-global, where $\delta = |F|/|A_n|$, then $\delta \leqslant \frac{\log^K n}{n}$.*

Remark 1.9. *A few remarks are in order.*

1. *We note that the above result achieves a stronger bound than the family in* (1). *There is no contradiction here, of course, since that family is very much not global: restricting to $\pi(1) = 2$ increases the measure of F significantly.*

2. *The junta method, which can be used to study many problems in extremal combinatorics, often considers the question for global families as a key component. The rough idea is to show that one can approximate a family F by a union of families $\tilde{F}$ that satisfy an appropriate pseudo-randomness condition, such that if F is product-free than so are the families $\tilde{F}$. Furthermore, inside any not-too-small pseudo-random family $\tilde{F}$, one may find a global family $\tilde{F}'$ by making any small restriction that increases the size of the family considerably. Thus, in this way one may hope to reduce the general question to the question on global families (see [18] for example).*

   While at the moment we do not know how to employ the junta method in the case of product-free sets in A_n, one may still hope that it is possible, providing some motivation for Theorem 1.8.

3. *Our result is in fact more general, and can be used to study the* 3*-set version of this problem; see Corollary 7.9.*

4. *We suspect that much stronger quantitative bounds should hold for global families; we elaborate on this suspicion in Section 7.2.4.*

1.4.3 Isoperimetric inequalities

Using our hypercontractive inequalities we are able to prove several isoperimetric inequalities for global sets. Let $S \subseteq S_n$ be a set, and consider the transpositions random walk T that from a permutation $\pi \in S_n$ moves to $\pi \circ \tau$, where τ is a randomly chosen transposition. We show that if S is "not too sensitive along any transposition",[6] then the probability to exit S in a random step according to T must be significant, similarly to the classical KKL Theorem on the hypercube [15]. The formal statement of this result is given in Theorem 7.13.

We are also able to analyze longer random walks according to T, of order $\approx n$, and show that one has small-set expansion for global sets. See Theorem 7.12 for a formal statement.

1.4.4 Deducing the results for other non-product domains

Our results for S_n imply analogous results in the multi-slice. The deduction is done in a black-box fashion, by a natural correspondence between functions over S_n and over the multi-slice that preserves degrees, globalness, and L_p norms.

This allows us to deduce analogs of our results for S_n essentially for free (see Section 7.4), as well as a stability result for the classical Kruskal–Katona Theorem (see Theorem 7.20).

1.4.5 Other applications

Our hypercontractive inequality has also been used in the study of Probabilistically Checkable Proofs [3]. More precisely, to study a new hardness conjecture, referred to as "Rich 2-to-1 Games Conjecture" in [3], and show that if true, it implies Khot's Unique-Games Conjecture [19].

1.5 Our techniques

In this section we outline the techniques used in the proofs of Theorem 1.2 and Theorem 1.4.

1.5.1 The coupling approach: proof overview

Obtaining hypercontractive operators via couplings

Consider two finite probability spaces X and Y, and suppose that $\mathcal{C} = (\mathbf{x}, \mathbf{y})$ is a coupling between them (we encourage the reader to think of X as S_n, and of Y as a product space in which we already know hypercontractivity to hold). Using the coupling $\mathcal{C}$, we may define the averaging operators $\mathrm{T}_{X\to Y}\colon L^2(X) \to L^2(Y)$ and $\mathrm{T}_{Y\to X}\colon L^2(Y) \to L^2(X)$ as

$$\mathrm{T}_{X\to Y} f(y) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{C}}\left[f(\mathbf{x}) \mid \mathbf{y} = y\right], \qquad \mathrm{T}_{Y\to X} f(x) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{C}}\left[f(\mathbf{y}) \mid \mathbf{x} = x\right].$$

It is easily noted by Jensen's inequality, that each one of the operators $\mathrm{T}_{X\to Y}$ and $\mathrm{T}_{Y\to X}$ is a contraction with respect to the L^p-norm, for any $p \geqslant 1$. The benefit of considering these operators, is that given an operator T_Y with desirable properties (say, it is hypercontractive, i.e. it satisfies $\|\mathrm{T}_Y f\|_4 \leqslant \|f\|_2$), we may consider the lifted operator on X given by $\mathrm{T}_X \stackrel{def}{=} \mathrm{T}_{Y\to X}\mathrm{T}_Y\mathrm{T}_{X\to Y}$ and hope that it too satisfies some desirable properties. Indeed, it is easy to see that if T_Y is hypercontractive, then T_X is also hypercontractive:

$$\|\mathrm{T}_{Y\to X}\mathrm{T}_Y\mathrm{T}_{X\to Y} f\|_4 \leqslant \|\mathrm{T}_Y\mathrm{T}_{X\to Y} f\|_4 \leqslant \|\mathrm{T}_{X\to Y} f\|_2 \leqslant \|f\|_2. \tag{2}$$

[6] The formal statement of the result requires an appropriate notion of discrete derivatives which we only give in Section 4.

We show that the same connection continues to hold for more refined hypercontractive inequalities such as the one given in [17, 18] (and more concretely, Theorem 2.5 below). We note that the proof in this case is slightly more involved.

While very elegant and appealing, the above approach can only be used to show hypercontractivity for a very special type of operators such as T_X defined above, and it is not clear if such results are of any use at all. To remedy this situation, we study the effect of this operator in the spectral domain. In particular, we show that the action of this operator on "low-degree functions" is very similar to the effect of the standard noise operator, and thus we are able deduce a hypercontractive inequality for low-degree functions, as in Theorem 1.4.

1.5.2 Instantiating the coupling approach for the symmetric group

The coupled space

Let $L = [n]^2$, and let m be large, depending polynomially on n ($m = n^2$ will do). We will couple S_n and L^m, where the idea is to think of each element of L as local information about the coupled permutation π. That is, the element $(i, j) \in L$ encodes the fact that π maps i to j.

Our coupling

We say that a set $T = \{(i_1, j_1), \ldots, (i_t, j_t)\} \subseteq L$ of pairs is *consistent* if there exists a permutation π with $\pi(i_k) = j_k$ for each $k \in [t]$, and any such permutation π is said to be consistent with T.

Our coupling between S_n and L^m is the following:

1. Choose an element $\mathbf{x} \sim L^m$ uniformly at random.
2. Greedily construct from $\mathbf{x}$ a set T of consistent pairs. That is, starting from $k = 1$ to m, we consider the k-th coordinate of $\mathbf{x}$, denoted by (i_k, j_k), and check whether adding it to T would keep it consistent. If so, we add (i_k, j_k) to T, and otherwise we do not.
3. Choose a permutation $\boldsymbol{\pi}$ consistent with T uniformly at random.

The resulting operator

Finally, we can specify our hypercontractive operator on S_n. Let $X = S_n$, $Y = L^m$ and $T_{X\to Y}, T_{Y\to X}$ be the operators corresponding to the coupling that we have just constructed. Let $\mathrm{T}_Y = \mathrm{T}_\rho$ be the noise operator on the product space L^m, which can be defined in two equivalent ways:

1. Every element is retained with probability ρ, and resampled otherwise.
2. The d'th Fourier level is multiplied by ρ^d.

Then $\mathrm{T}^{(\rho)} = \mathrm{T}_{Y\to X}\mathrm{T}_Y\mathrm{T}_{X\to Y}$ is our desired operator on S_n.

We next explain how to analyze the operator T_Y.

Showing that T_Y satisfies refined hypercontractivity

Recall the simplistic argument (2), showing that hypercontractivity of T_X implies that the hypercontractivity of T_Y. We intend to show, in a similar way, that refined hypercontractivity is also carried over by the coupling. Towards this end, we must show that the notion of globalness is preserved: namely, if f is global, then $g = \mathrm{T}_{S_n \to L^m} f$ is also global. This assertion however very much depends on the precise notion of globalness we consider. If we assume that f is ε-global with constant C, then it is easy to show that g is also ε-global with constant C (see Proposition 3.1), and the argument goes through smoothly. However, in the case that f is only guaranteed to be (d, ε)-global, things are more interesting, and in this case we are only able to handle f's that are of low-degree (this is natural, as we will deal with the low-degree part of (d, ε)-global functions).

A convenient feature of product spaces is that for low-degree functions, the notions of ε-globalness with constant C, and (D, δ)-globalness, are equivalent up to small losses in parameters. This allows one to invoke results such as Theorem 2.5 in this case. While we show that the case of the symmetric group possesses a similar property (at least when n is large enough in comparison to d), we are not able to immediately use it. The issue is that even if $f\colon S_n \to \mathbb{R}$ is a function of degree d, it may not be the case that $g = \mathrm{T}_{S_n \to L_m} f$ is also of low degree.

We circumvent this issue as follows. Suppose f is $(2d, \varepsilon)$-global is of degree d. Then, as remarked above, we argue that f is ε-global with some absolute constant C, and so it is $(t, C^t\varepsilon)$-global for all $t \in \mathbb{N}$. Thus, g is $(t, C^t\varepsilon)$-global for all t. Now, as g is a function over a product space, it is easily seen that the latter implies that the noisy version of g, $h = \mathrm{T}_{\frac{1}{4C}} g$, is ε-global with factor 2, and thus we are able to invoke Theorem 2.5 on it. Together, this implies that taking $\mathrm{T}_Y = \mathrm{T}_{1/80^2} \circ \mathrm{T}_{1/(4C)}$ gets us that $\|\mathrm{T}_X f\|_4 \leqslant \sqrt{\varepsilon \|f\|_2}$. (The constant $1/80^2$ arises from Theorem 2.5.)

1.5.3 The direct approach: proof overview

Our second approach to establish hypercontractive inequalities goes via a rather different route. One of the proofs of hypercontractivity in product domains proceeds by finding a convenient, orthonormal basis for the space of real-valued functions over Ω (which in product cases is easy as the basis tensorizes). This way, proving hypercontractivity amounts to studying moments of this basis functions as well as other forms, which is often not very hard to do due to the simple nature of the basis.

When dealing with non-product spaces, such as S_n, we do not know how to produce such a convenient orthonormal basis. Nevertheless, our direct approach presented in Section 6 relies on a representation of a function $f\colon S_n \to \mathbb{R}$ in a canonical form that is almost as good as in product spaces. To construct this representation, we start with obvious spanning sets such as

$$\left\{ \prod_{\ell=1}^{d} 1_{\pi(i_\ell)=j_\ell} \,\middle|\, |\{i_1, \ldots, i_d\}| = |\{j_1, \ldots, j_d\}| = d \right\}.$$

This set contains many redundancies (and thus is not a basis), and we show how to use these to enforce a system of linear constraints on the coefficients of the representation that turn out to be very useful in proving hypercontractive inequalities.

1.6 Organization of the paper

In Section 2 we present some basic preliminaries. Sections 3, 4 and 5 are devoted for presenting our approach to hypercontractivity via coupling and algebraic arguments, and in Section 6 we present our direct approach. In Sections 7 and 8 we present several consequences of our hypercontractive inequalities: the level-d inequality in Section 8, and the other applications in Section 7.

2 Preliminaries

We think of the product operation in S_n as function composition, and so $(\tau\sigma)(i) = (\tau \circ \sigma)(i) = \tau(\sigma(i))$.

Throughout the paper, we consider the space of real-valued functions on S_n equipped with the expectation inner product, denoted by $L^2(S_n)$. Namely, for any $f, g\colon S_n \to \mathbb{R}$ we define $\langle f, g\rangle = \mathbb{E}_{\sigma\in S_n}[f(\sigma)g(\sigma)]$. A basic property of this space is that it is an S_n-bimodule, as can be seen by defining the left operation on a function f and a permutation τ as ${}^{\tau}f(\sigma) = f(\tau \circ \sigma)$, and the right operation $f^{\tau}(\sigma) = f(\sigma \circ \tau)$.

2.1 The level decomposition

We will define the concept of *degree d function* in several equivalent ways. The most standard definition is the one which we already mentioned in the introduction.

Definition 2.1. *Let $T = \{(i_1, j_1), \dots, (i_t, j_t)\} \subseteq L$ be a set of t consistent pairs, and recall that S_n^T is the set of all permutations such that $\pi(i_k) = j_k$ for all $k \in [t]$.*

The space V_d consists of all linear combinations of functions of the form $1_T = 1_{S_n^T}$ for $|T| \leqslant d$. We say that a real-valued function on S_n has degree (at most) d if it belongs to V_d.

By construction, $V_{d-1} \subseteq V_d$ for all $d \geqslant 1$. We define the space of functions of pure degree d as

$$V_{=d} = V_d \cap V_{d-1}^{\perp}.$$

It is easy to see that $V_n = V_{n-1}$, and so we can decompose the space of all real-valued functions on S_n as follows:

$$\mathbb{R}[S_n] = V_{=0} \oplus V_{=1} \oplus \cdots \oplus V_{=n-1}.$$

We comment that the representation theory of S_n refines this decomposition into a finer one, indexed by partitions λ of n; the space $V_{=d}$ corresponds to partitions in which the largest part is exactly $n - d$.

We may write any function $f\colon S_n \to \mathbb{R}$ in terms of our decomposition uniquely as $\sum\limits_{i=0}^{n-1} f^{=i}$, where $f^{=i} \in V_{=i}$. It will also be convenient for us to have a notation for the projection of f onto V_d, which is nothing but $f^{\leqslant d} = f^{=0} + f^{=1} + \cdots + f^{=d}$.

We will need an alternative description of $V_{=d}$ in terms of juntas.

Definition 2.2. *Let $A, B \subseteq [n]$. For every $a \in A$ and $b \in B$, let $e_{ab} = 1_{\pi(a)=b}$. We say that a function $f\colon S_n \to \mathbb{R}$ is an (A, B)-junta if f can be written as a function of the e_{ab}. We denote the space of (A, B)-juntas by $V_{A,B}$.*

A function is a d-junta if it is an (A, B)-junta for some $|A| = |B| = d$.

Lemma 2.3. *The space $V_{A,B}$ is spanned by the functions 1_T for $T \subseteq A \times B$. Consequently, V_d is the span of the d-juntas.*

Proof. If $A = \{i_1, \dots, i_d\}$ and $B = \{j_1, \dots, j_d\}$ then an (A, B)-junta f can be written as a function of $e_{i_s j_t}$, and in particular as a polynomial in these functions. Since $e_{i_s j_{t_1}} e_{i_s j_{t_2}} = e_{i_{s_1} j_t} e_{i_{s_2} j_t} = 0$ if $t_1 \neq t_2$ and $s_1 \neq s_2$, it follows that f can be written as a linear combination of functions 1_T for $T \subseteq A \times B$.

Conversely, if $T = \{(a_1, b_1), \dots, (a_d, b_d)\}$ then $1_T = e_{a_1 b_1} \cdots e_{a_d b_d}$.

To see the truth of the second part of the lemma, notice that if $|A| = |B| = d$ and $T \subseteq A \times B$ then $|T| \leqslant d$, and conversely if $|T| \leqslant d$ then $T \subseteq A \times B$ for some A, B such that $|A| = |B| = d$. $\square$

We will also need an alternative description of $V_{A,B}$.

Lemma 2.4. *For each A, B, the space $V_{A,B}$ consists of all functions $f\colon S_n \to \mathbb{R}$ such that $f = {}^{\tau}f^{\sigma}$ for all σ fixing A pointwise and τ fixing B pointwise.*

Proof. Let $U_{A,B}$ consist of all functions f satisfying the stated condition, i.e., $f(\pi) = f(\tau\pi\sigma)$ whenever σ fixes A pointwise and τ fixes B pointwise.

Let $a \in A$ and $b \in B$. If σ fixes a and τ fixes b then $\pi(a) = b$ iff $\tau\pi\sigma(a) = b$, showing that $e_{ab} \in U_{A,B}$. It follows that $V_{A,B} \subseteq U_{A,B}$.

In the other direction, let $f \in U_{A,B}$. Suppose for definiteness that $A = [a]$ and $B = [b]$. Let π be a permutation such that $\pi(1) = 1, \dots, \pi(t) = t$, and $\pi(i) > b$ for $i = t+1, \dots, a$. Applying a permutation fixing B pointwise on the left, we turn π into a permutation π' such that $\pi'(1), \dots, \pi'(a) = 1, \dots, t, b+1, \dots, b+(a-t)$. Applying a permutation fixing A pointwise on the right, we turn π' into the permutation $1, \dots, t, b+1, \dots, b+(a-t), \dots, n, t+1, \dots, a$. This shows that if π_1, π_2 are two permutations satisfying $e_{ab}(\pi_1) = e_{ab}(\pi_2)$ for all $a \in A, b \in B$ then we can find permutations σ_1, σ_2 fixing A pointwise and permutations τ_1, τ_2 fixing B pointwise such that $\tau_1\pi_1\sigma_1 = \tau_2\pi_2\sigma_2$, and so $f(\pi_1) = f(\pi_2)$. This shows that $f \in V_{A,B}$. $\square$

2.2 Hypercontractivity in product spaces

We will make use of the following hypercontractive inequality, essentially due to [18]. For that, we first remark that we consider the natural analog definitions of globalness for product spaces. Namely, for a finite product space $(\Omega, \mu) = (\Omega_1 \times \dots \times \Omega_m, \mu_1 \times \dots \times \mu_m)$, we say that $f\colon \Omega \to \mathbb{R}$ is ε-global with a constant C, if for any $T \subseteq [m]$ and $x \in \prod_{i \in T} \Omega_i$ it holds that $\|f_{T \to x}\|_{2,\mu_x}^2 \leqslant C^{|T|}\varepsilon$, where μ_x is the distribution μ conditioned on coordinates of T being equal to x. Similarly, we say that f is (d, ε)-global if for any $|T| \leqslant d$ and $x \in \prod_{i \in T} \Omega_i$ it holds that $\|f_{T \to x}\|_{2,\mu_x}^2 \leqslant \varepsilon$.

Theorem 2.5. *Let $q \in \mathbb{N}$ be even, and suppose f is ε-global with constant C, and let $\rho \leqslant \frac{1}{(10qC)^2}$. Then $\|\mathrm{T}_\rho f\|_q \leqslant \varepsilon^{q-2} \|f\|_2^{\frac{2}{q}}$.*

We remark that Theorem 2.5 was proved in [18] for $q = 4$, however the proof is essentially the same for all even integers q.

3 Hypercontractivity: the coupling approach

3.1 Hypercontractivity from full globalness

In this section we prove the following hypercontractive results for our operator $\mathrm{T}^{(\rho)}$ assuming f is global. We begin by proving two simple propositions.

Proposition 3.1. *Suppose $f\colon S_n \to \mathbb{R}$ is ε-global with constant C, and let $g = \mathrm{T}_{S_n \to L^m} f$. Then g is ε-global with constant C.*

Proof. Let S be a set of size t, and let $x = \left((i_k, j_k)\right)_{k\in S} \in L^S$. Let $y \sim L^{[m]\setminus S}$ be chosen uniformly, and let σ be the random permutation that our coupling process outputs given (x, y). We have

$$\|g_{S\to x}\|_2^2 = \mathbb{E}_y \left(\mathbb{E}_\sigma f\left(\sigma\right)\right)^2 \leqslant \mathbb{E}_\sigma \left[f\left(\sigma\right)^2\right]$$

by Cauchy–Schwarz. Next, we consider the values of $\sigma\left(i_k\right)$ for $k \in S$, condition on them and denote $T = \{(i_k, \sigma(i_k))\}$. The conditional distribution of σ given T is uniform by the symmetry of elements in $[n] \setminus \{i_k | \, k \in S\}$, so for any permutation π on $[n] \setminus \{i_k | \, k \in S\}$ we have that $\sigma\pi$ has the same probability as σ. Also, the collection $\{\sigma\pi\}$ consists of all permutations satisfying T, so

$$\mathbb{E}\left[f\left(\sigma\right)^2\right] = \mathbb{E}_T\left[\|f_T\|_2^2\right] \leqslant \max_T \|f_T\|_2^2 \leqslant C^{2|S|}\varepsilon^2. \qquad \square$$

Fact 3.2. *Suppose that we are given two probability spaces $(X, \mu_X), (Y, \mu_Y)$. Suppose further that for each $x \in X$ we have a distribution $N(x)$ on Y, such that if we choose $x \sim \mu_X$ and $y \sim N(x)$, then the marginal distribution of y is μ_Y. Define an operator $\mathrm{T}_{Y\to X} \colon L^2(Y) \to L^2(X)$ by setting*

$$\mathrm{T}_{Y\to X} f\left(x\right) = \mathbb{E}_{y\sim N(x)} f\left(y\right).$$

Then $\|\mathrm{T}_{Y\to X} f\|_q \leqslant \|f\|_q$ for each $q \geqslant 1$.

We can now prove one variant of our hypercontractive inequality for global functions over the symmetric group.

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

Proof. Let $f \colon S_n \to \mathbb{R}$ be ε-global with constant C. By Proposition 3.1, the function $g = \mathrm{T}_{S_n \to L^m} f$ is also ε-constant with constant C, and by Fact 3.2 we have

$$\left\|\mathrm{T}^{(\rho)} f\right\|_q^q = \|\mathrm{T}_{L^m\to S_n} \mathrm{T}_\rho g\|_q^q \leqslant \|\mathrm{T}_\rho g\|_q^q.$$

Now, by Theorem 2.5 we may upper-bound the last norm by $\varepsilon^{q-2}\|g\|_2^2$, and using Fact 3.2 again we may bound $\|g\|_2^2 \leqslant \|f\|_2^2$. $\square$

Remark 3.4. *Once the statement has been proven for even q's, a qualitatively similar statement can be automatically deduced for all q's, as follows. Fix q, and take the smallest $q \leqslant q' \leqslant q+2$ that is an even integer. Then for $\rho \leqslant \frac{1}{(10(q+2)C)^2} \leqslant \frac{1}{(10q'C)^2}$ we may bound*

$$\left\|\mathrm{T}^{(\rho)} f\right\|_q \leqslant \left\|\mathrm{T}^{(\rho)} f\right\|_{q'} \leqslant \varepsilon^{\frac{q'-2}{q'}} \|f\|_2^{\frac{2}{q'}} \leqslant \varepsilon^{\frac{q}{q+2}} \|f\|_2^{\frac{2}{q+2}},$$

where in the last inequality we used $q' \leqslant q+2$ and $\|f\|_2 \leqslant \varepsilon$.

3.2 Hypercontractivity for low-degree functions

Next, we use Theorem 3.3 to prove our hypercontractive inequality for low-degree functions that assumes considerably weaker globalness properties of f, namely Theorem 1.4. The proof of the above theorem makes use of the following key lemmas. The first of which asserts that just like in the cube, bounded globalness of a low-degree function implies (full) globalness.

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

Thus, to deduce Theorem 1.4 from Theorem 3.3, it suffices to show that f may be approximated by linear combinations of $\mathrm{T}^{(\rho^i)}f$ for $i = 1, 2, \ldots$ in L^q, and this is the content of our second lemma. First, let us introduce some convenient notations. For a polynomial $P(z) = a_0 + a_1 z + \cdots + a_k z^k$, we denote the spectral norm of P by $\|P\| = \sum_{i=0}^{k} |a_i|$. We remark that it is easily seen that $\|P_1 P_2\| \leqslant \|P_1\|\|P_2\|$ for any two polynomials P_1, P_2.

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

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

for every function f of degree at most d.

We defer the proofs of Lemmas 3.5 and 3.6 to Sections 4 and 5, respectively. In the remainder of this section we derive Theorem 1.4 from them, restated below.

Theorem 1.4 (Restated) . *There exists $C > 0$ such that the following holds. Let $q \in \mathbb{N}$ be even, $n \geqslant q^{C\cdot d^2}$. If f is a $(2d,\varepsilon)$-global function of degree d, then $\|f\|_q \leqslant q^{O\left(d^3\right)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$.*

Proof. Choose $\rho = 1/(400C^3q^2)$, and let P be as in Lemma 3.6. Then

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

As for the first term, we have

$$\left\| \sum_{i=1}^{l} a_i \left(\mathrm{T}^{(\rho)}\right)^i f \right\|_q \leqslant \sum_{i=1}^{l} |a_i| \left\| \left(\mathrm{T}^{(\rho)}\right)^i f \right\|_q \leqslant \|P\| \left\| \mathrm{T}^{(\rho)} f \right\|_q \leqslant q^{O\left(d^3\right)} \left\| \mathrm{T}^{(\rho)} f \right\|_q .$$

To estimate $\left\|\mathrm{T}^{(\rho)} f\right\|_q$, note first that by Lemma 3.5, f is ε-global for constant 4^8, thus given that C is large enough we may apply Theorem 3.3 to deduce that $\left\|\mathrm{T}^{(\rho)} f\right\|_q \leqslant \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}$. As $\|f\|_2 \leqslant \varepsilon$ we conclude that

$$\|f\|_q \leqslant q^{O\left(d^3\right)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}} + \frac{1}{\sqrt{n}} \|f\|_2 = q^{O\left(d^3\right)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}} . \qquad \square$$

4 Proof of Lemma 3.5

We begin by proving Lemma 3.5. A proof of the corresponding statement in product spaces proceeds by showing that a function is (d, ε)-global if and only if the 2-norms of derivatives of f of order d are small. Since then derivatives of order higher than d of f are automatically 0 (by degree considerations), they are automatically small. Thus, if f is a (d, ε)-global function of degree d, then *all derivatives* of f have small 2-norm, and by the reverse relation it follows that f is ε-global for some constant C.

Our proof follows a similar high level idea. The main challenge in the proof is to find an appropriate analog of discrete derivatives from product spaces, that both reduces the degree of the function f and can be related to restrictions of f. Towards this end, we make the following key definition.

Definition 4.1. *Let $i_1 \neq i_2 \in [n]$ and $j_1 \neq j_2 \in [n]$.*

1. *The Laplacian of f along (i_1, i_2) is defined as $\mathrm{L}_{(i_1,i_2)}[f] = f - f^{(i_1\, i_2)}$, where we denote by $(i_1\, i_2)$ the transposition of i_1 and i_2.*

2. *The derivative of f along $(i_1, i_2) \to (j_1, j_2)$ is $(\mathrm{L}_{(i_1,i_2)} f)_{(i_1,i_2)\to(j_1,j_2)}$. More explicity, it is a function defined on $S_n^{(i_1,j_1),(i_2,j_2)}$ (that is isomorphic to S_{n-2}) whose value on π is*

$$f(\pi) - f(\pi \circ (i_1, i_2)).$$

3. *For distinct $i_1, \ldots, i_t$ and distinct $j_1, \ldots, j_t$, denote the ordered set $S = \{(i_1, j_1), \ldots, (i_t, j_t)\}$ and define the Laplacian of f along S as $L_S[f] = L_{i_1,j_1} \circ \cdots \circ L_{i_t,j_t} \circ f$.*

   For $(k_1, \ell_1), \ldots, (k_t, \ell_t)$, the derivative of f along $S \to \{(k_1, \ell_1), \ldots, (k_t, \ell_t)\}$ is

$$\mathrm{D}_{S\to\{(k_1,l_1),\ldots,(k_t,l_t)\}} f = (L_S[f])_{S\to\{(i_1,k_1),(j_1,l_1),\ldots,(i_t,k_t),(j_t,l_t)\}}$$

   We call D a derivative of order t. We also include the case where $t = 0$, and call the identity operator a 0-derivative.

The following two claims show that the definition of derivatives above is good, in the sense that 2-norms of derivatives relate to globalness, and derivatives indeed reduce the degree of f.

Claim 4.2. *Let $t \in \mathbb{N}$, and $\varepsilon > 0$, and $f : S_n \to \mathbb{R}$.*

1. *If f is $(2t, \varepsilon)$-global, then for each derivative D of order t we have that $\|\mathrm{D}f\|_2 \leqslant 2^t \varepsilon$.*

2. *If $t \leqslant n/2$, and for all $\ell \leqslant t$ and every derivative D of order ℓ we have that $\|\mathrm{D}f\|_2 \leqslant \varepsilon$, then f is $(t, 2^t\varepsilon)$-global.*

Proof. The first item follows immediately by induction on t using the triangle inequality. The rest of the proof is devoted to establishing the second item, also by induction on t.

Base case $t = 0, 1$. The case $t = 0$ is trivial, and we prove the case $t = 1$. Let $i_1, i_2 \in [n]$ be distinct and let $j_1, j_2 \in [n]$ be distinct. Since $\|\mathrm{D}_{(i_1,i_2)\to(j_1,j_2)} f\|_2 \leqslant \varepsilon$ we get from the triangle inequality that

$$\left| \|f_{i_1\to j_1, i_2\to j_2}\|_2 - \|f_{i_2\to j_1, i_1\to j_2}\|_2 \right| \leqslant \varepsilon. \tag{3}$$

Multiplying (3) by $\|f_{i_1\to j_1, i_2\to j_2}\|_2 + \|f_{i_2\to j_1, i_1\to j_2}\|_2$ we get that

$$\left|\|f_{i_1\to j_1,i_2\to j_2}\|_2^2 - \|f_{i_2\to j_1,i_1\to j_2}\|_2^2\right| \leqslant \varepsilon\left(\|f_{i_1\to j_1,i_2\to j_2}\|_2 + \|f_{i_2\to j_1,i_1\to j_2}\|_2\right).$$

Taking average over j_2 and using the triangle inequality on the left-hand side, we get that

$$\left|\|f_{i_1\to j_1}\|_2^2 - \|f_{i_2\to j_1}\|_2^2\right| \leqslant \varepsilon \mathbb{E}_{j_2}\left[\|f_{i_1\to j_1,i_2\to j_2}\|_2 + \|f_{i_2\to j_1,i_1\to j_2}\|_2\right].$$

By Cauchy–Schwarz, $\mathbb{E}_{j_2}\left[\|f_{i_1\to j_1,i_2\to j_2}\|_2\right] \leqslant \mathbb{E}_{j_2}\left[\|f_{i_1\to j_1,i_2\to j_2}\|_2^2\right]^{1/2} = \|f_{i_1\to j_1}\|_2$, and similarly for the other term, so we conclude

$$\left|\|f_{i_1\to j_1}\|_2^2 - \|f_{i_2\to j_1}\|_2^2\right| \leqslant \varepsilon\left(\|f_{i_1\to j_1}\|_2 + \|f_{i_2\to j_1}\|_2\right),$$

and dividing both sides of the inequality by $\|f_{i_1\to j_1}\|_2 + \|f_{i_2\to j_1}\|_2$ we get

$$\left|\|f_{i_1\to j_1}\|_2 - \|f_{i_2\to j_1}\|_2\right| \leqslant \varepsilon.$$

Since $\mathbb{E}_{i_2\sim[n]}\|f_{i_2\to j_1}\|_2^2 = \|f\|_2^2 \leqslant \varepsilon$, we get that there is i_2 such that $\|f_{i_2\to j_1}\|_2 \leqslant \varepsilon$, and the above inequality implies that $\|f_{i_1\to j_1}\|_2 \leqslant 2\varepsilon$ for all i_1. This completes the proof for the case $t = 1$.

The inductive step. Let $t > 1$. We prove that f is $\left(t, 2^t\varepsilon\right)$-global, or equivalently that f_T is $\left(1, 2^t\varepsilon\right)$-global for all consistent sets T of size $t-1$. Indeed, fix a consistent T of size $t-1$.

By the induction hypothesis, $\|f_T\|_2 \leqslant 2^{t-1}\varepsilon$, and the claim would follow from the $t = 1$ case once we show that $\|\mathrm{D}f_{\to T}\|_2 \leqslant 2^{t-1}\varepsilon$ for all order 1 derivatives $\mathrm{D} = \mathrm{D}_{(i_1,i_2)\to(j_1,j_2)}$, where i_1, i_2 do not appear as the first coordinate of an element in T, and j_1, j_2 do not appear as a second coordinate of an element of T (we're using the fact here that the case $t = 1$ applies, as S_n^T is isomorphic to $S_{n-|T|}$ as $S_{n-|T|}$-bimodules). Fix such D, and let $g = \mathrm{D}_{(i_1,i_2)\to(j_1,j_2)}f$. By hypothesis, for any order $t-1$ derivative $\tilde{\mathrm{D}}$ we have that $\|\tilde{\mathrm{D}}g\|_2 \leqslant \varepsilon$, hence by the induction hypothesis $\|g_{\to T}\|_2 \leqslant 2^{t-1}\varepsilon$. Since restrictions and derivatives commute, we have $g_{\to T} = \mathrm{D}_{(i_1,i_2)\to(j_1,j_2)}f_{\to T}$, and we conclude that $f_{\to T}$ is $\left(1, 2^t\varepsilon\right)$-global, as desired. $\square$

Claim 4.3. *If f is of degree d, and D is a t-derivative, then $\mathrm{D}f$ is of degree $\leqslant d - t$.*

Proof. It is sufficient to consider the case $t = 1$ of the proposition, as we may apply it repeatedly. By linearity of the derivative D it is enough to show it in the case where $f = x_{i_1\to j_1}\cdots x_{i_t\to j_t}$. Now note that the Laplacian $L_{(k_1 k_2)}$ annihilates f unless either k_1 is equal to some i_ℓ, or k_2 is equal to some i_ℓ, or both, and we only have to consider these cases. Each derivative corresponding to the Laplacian $L_{(k_1,k_2)}$ restricts both the image of k_1 and the image of k_2, so after applying this restriction on $L_{(k_1,k_2)}f$ we either get the 0 function, a function of degree $d-1$, or a function of degree $d-2$. $\square$

We are now ready to prove Lemma 3.5. To prove that f is global, we handle restrictions of size $t \leqslant n/2$, and restrictions of size $t > n/2$ separately, in the following two claims.

Claim 4.4. *Suppose $f\colon S_n \to \mathbb{R}$ is a $(2d, \varepsilon)$-global function of degree d. Then f is $\left(t, 4^t\varepsilon\right)$-global for each $t \leqslant \frac{n}{2}$.*

Proof. By the second item in Claim 4.2, it is enough to show that for each t-derivative D we have $\|\mathrm{D}f\|_2 \leqslant 2^t\varepsilon$. For $t \leqslant d$ this follows from the first item in Claim 4.2, and for $t > d$ it follows from Proposition 4.3 as we have that $\mathrm{D}f = 0$ for all derivatives of order t. $\square$

For $t \geqslant \frac{n}{2}$, we use the obvious fact f is always $(t, \|f\|_\infty)$-global, and upper bound the infinity norm of f using the following claim.

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

Proof. We prove the claim by induction on n. The case $n = 1$ is obvious, so let $n > 1$.

If $3d \leqslant \frac{n}{2}$, then by Claim 4.4 we have that f is $\left(3d, 4^{3d}\varepsilon\right)$-global, and hence for each set S of size d, the function $f_{\to S}$ is $\left(2d, 4^{3d}\varepsilon\right)$-global. Therefore, the induction hypothesis implies that

$$\|f\|_\infty = \max_{S\colon |S|=d} \|f_S\|_\infty \leqslant \sqrt{(6d)!}4^{3(n-d)} \cdot 4^{3d}\varepsilon = \sqrt{(6d)!}4^{3n}\varepsilon.$$

Suppose now that $n \leqslant 6d$. Then $\|f\|_\infty^2 \leqslant (6d)!\|f\|_2^2$ since the probability of each atom in S_{6d} is $\frac{1}{(6d)!}$. Hence, $\|f\|_\infty \leqslant \sqrt{(6d)!}\varepsilon$. □

Note that $(6d)! \leqslant 4^n$ given C is sufficiently large, so for $t > n/2$, Claim 4.5 implies that f is $(t, 4^{4n}\varepsilon) = (t, 4^{8t}\varepsilon)$-global. □

5 Proof of Lemma 3.6

Proof overview. Our argument first constructs a very strong approximating polynomial in the L_2-norm. The approximation will be in fact strong enough to imply, in a black-box way, that it is also an approximating polynomial in L_q.

To construct an L_2 approximating polynomial, we use spectral considerations. Denote by $\lambda_1, \ldots, \lambda_\ell$ the eigenvalues of $\mathrm{T}^{(\rho)}$ on the space of degree d functions. Note that if P is a polynomial such that $P(\lambda_i) = 1$ for all i, then $P(\mathrm{T}^\rho)f = f$ for all f of degree d. However, as ℓ may be very large, there may not be a polynomial P with small $\|P\|$ satisfying $P(\lambda_i) = 1$ for all i, and to circumvent this issue we must argue that, at least effectively, ℓ is small. Indeed, while we do not show that ℓ is small, we do show that there are d distinct values, $\lambda_1(\rho), \ldots, \lambda_d(\rho)$, such that each λ_i is very close to one of the $\lambda_j(\rho)$'s. This, by interpolation, implies that we may find a low-degree polynomial P such that $P(\lambda_i)$ is very close to 1 for all $i = 1, \ldots, \ell$. Finally, to argue that $\|P\|$ is small, we show that each $\lambda_i(\rho)$ is bounded away from 0.

It remains then to establish the claimed properties of the eigenvalues $\lambda_1, \ldots, \lambda_\ell$, and we do so in several steps. We first identify the eigenspaces of $\mathrm{T}^{(\rho)}$ among the space of low-degree functions, and show that each one of them contains a junta. Intuitively, for juntas it is much easier to understand the action of the $\mathrm{T}^{(\rho)}$, since when looking on very few coordinates, S_n looks like a product space. Indeed, using this logic we are able to show that all eigenvalues of $\mathrm{T}^{(\rho)}$ on low-degree functions are bounded away from 0. To argue that the eigenvalues are concentrated on a few values, we use the fact that taking symmetry into account, the number of linearly independent juntas is small.

Our proof uses several notations appearing in Section 2.1, including the actions of S_n on functions from the left ${}^\tau f$ and from the right f^σ, the level decomposition V_d, the spaces $V_{A,B}$, and the concept of d-junta.

5.1 Identifying the eigenspaces of $\mathrm{T}^{(\rho)}$

5.1.1 $\mathrm{T}^{(\rho)}$ commutes with the action of S_n as a bimodule

Lemma 5.1. *The operator $\mathrm{T}^{(\rho)}$ commutes with the action of S_n as a bimodule.*

The proof relies on the following claims.

Claim 5.2. *If* T, S *are operators that commute with the action of* S_n *as a bimodule, then so is* $\mathrm{T} \circ \mathrm{S}$.

Proof. We have ${}^{\pi_1}(\mathrm{T}Sf)^{\pi_2} = \mathrm{T}\left({}^{\pi_1}Sf^{\pi_2}\right) = \mathrm{TS}({}^{\pi_1}f^{\pi_2})$. $\square$

Let X and Y be S_n-bimodules, and consider $X \times Y$ as an S_n-bimodule with the operation ${}^{\sigma_1}(x, y)^{\sigma_2} = ({}^{\sigma_1}x^{\sigma_2}, {}^{\sigma_1}y^{\sigma_2})$. We say that a probability distribution μ on $X \times Y$ is invariant under the action of S_n on both sides if $\mu({}^{\sigma_1}(x, y)^{\sigma_2}) = \mu(x, y)$ for all $x \in X$, $y \in Y$ and $\sigma_1, \sigma_2 \in S_n$.

Claim 5.3. *Let* X, Y *be* S_n*-bimodules that are coupled by the probability measure* μ*, and suppose that* μ *is invariant under the action of* S_n *from both sides. Then the operators* $\mathrm{T}_{X\to Y}, \mathrm{T}_{Y\to X}$ *commute with the action of* S_n *from both sides.*

Proof. We prove the claim for $\mathrm{T}_{X\to Y}$ (the argument for $\mathrm{T}_{Y\to X}$ is identical). Let μ_X, μ_Y be the marginal distributions of μ on X and on Y, and for each $x \in X$ denote by 1_x the indicator function of x. Then the set $\{1_x\}_{x\in X}$ is a basis for $L^2(X)$, and so it is enough to show that for all x and $\sigma_1, \sigma_2 \in S_n$ it holds that ${}^{\sigma_1}(\mathrm{T}_{X\to Y}1_x)^{\sigma_2} = \mathrm{T}_{X\to Y}\left({}^{\sigma_1}1_x{}^{\sigma_2}\right)$. Note that as these are two functions over Y, it is enough to show that

$$\langle {}^{\sigma_1}(\mathrm{T}_{X\to Y}1_x)^{\sigma_2}, 1_y\rangle = \langle \mathrm{T}_{X\to Y}\left({}^{\sigma_1}1_x{}^{\sigma_2}\right), 1_y\rangle$$

for all y, since $\{1_y\}_{y\in Y}$ forms a basis for $L^2(Y)$.

Fix x and y. Since μ is invariant under the action of S_n on both sides, it follows that μ_Y is invariant under the action of S_n, so we have

$$\langle {}^{\sigma_1}(\mathrm{T}_{X\to Y}1_x)^{\sigma_2}, 1_y\rangle = \left\langle \mathrm{T}_{X\to Y}1_x, {}^{\sigma_1^{-1}}1_y{}^{\sigma_2^{-1}}\right\rangle = \left\langle \mathrm{T}_{X\to Y}1_x, 1_{\sigma_1 y\sigma_2}\right\rangle = \mu\left(x, \sigma_1 y\sigma_2\right),$$

where in the penultimate transition we used the fact that ${}^{\sigma_1^{-1}}1_y{}^{\sigma_2^{-1}} = 1_{\sigma_1 y^{\sigma_2}}$. On the other hand, we also have that the last fact holds for 1_x, and so

$$\langle \mathrm{T}_{X\to Y}\left({}^{\sigma_1}1_x{}^{\sigma_2}\right), 1_y\rangle = \left\langle \mathrm{T}_{X\to Y}1_{\sigma_1^{-1}x^{\sigma_2^{-1}}}, 1_y\right\rangle = \mu\left(\sigma_1^{-1}x\sigma_2^{-1}, y\right).$$

The claim now follows from the fact that μ is invariant under the action of S_n from both sides. $\square$

We are now ready to move on to the proof of Lemma 5.1.

Proof of Lemma 5.1. We let S_n act on L from the right by setting $(i, j)\,\pi = (\pi(i), j)$ and from the left by setting $\pi(i, j) = (i, \pi(j))$. For a function f on L^m we write ${}^{\pi_1}f^{\pi_2}$ for the function

$$(x_1, \ldots, x_m) \mapsto f\left(\pi_1 x_1 \pi_2, \ldots, \pi_1 x_m \pi_2\right).$$

By Claim 5.3 the operators $\mathrm{T}_\rho, \mathrm{T}_{S_n\to L^m}, \mathrm{T}_{L^m\to S_n}$ commute with the action of S_n as a bimodule, and therefore so is $\mathrm{T}^{(\rho)}$ by Claim 5.2. $\square$

5.1.2 Showing that the spaces $V_{A,B}$ and V_d are invariant under $\mathrm{T}^{(\rho)}$

First we show that $V_{A,B}$ is an invariant subspace of $\mathrm{T}^{(\rho)}$.

Lemma 5.4. *Let* T *be an endomorphism of* $L^2(S_n)$ *as an* S_n*-bimodule. Then* $\mathrm{T}V_{A,B} \subseteq V_{A,B}$*. Moreover,* $TV_d \subseteq V_d$*.*

Proof. Let $f \in V_{A,B}$. We need to show that $\mathrm{T}f \in V_{A,B}$. Let $\sigma_1 \in S_{[n]\setminus A}, \sigma_2 \in S_{[n]\setminus B}$. Then

$$^{\sigma_1}(\mathrm{T}f)^{\sigma_2} = \mathrm{T}\left(^{\sigma_1}f^{\sigma_2}\right) = \mathrm{T}f,$$

where the first equality used the fact that T commutes with the action of S_n from both sides, and the second inequality follows from Lemma 2.4. The 'moreover' part follows from Lemma 2.3. □

Lemma 5.5. *Let λ be an eigenvalue of $\mathrm{T}^{(\rho)}$ as an operator from V_d to itself. Let $V_{d,\lambda}$ be the eigenspace corresponding to λ. Then $V_{d,\lambda}$ contains a d-junta.*

Proof. Since each space $V_{A,B}$ is $\mathrm{T}^{(\rho)}$ invariant, we may decompose each $V_{A,B}$ into eigenspaces $V_{A,B}^{(\lambda)}$. Let

$$V_d^{(\lambda)} = \sum_{|A|,|B|\leqslant d} V_{A,B}^{(\lambda)}.$$

Then for each λ, $V_d^{(\lambda)}$ is an eigenspaces of $\mathrm{T}^{(\rho)}$ with eigenvalue λ, and

$$\sum_{\lambda} V_d^{(\lambda)} = \sum_{|A|,|B|\leqslant d} V_{A,B} = V_d = \sum_{\lambda} V_{d,\lambda}.$$

By uniqueness, it follows that $V_{d,\lambda} = V_d^{(\lambda)}$ for all λ. Fix λ; then we get that there are $|A|,|B| \leqslant d$ such that $V_{A,B}^{\lambda} \subseteq V_{d,\lambda}$, and since any function in $V_{A,B}$ is a d-junta by definition, the proof is concluded. □

We comment that the representation theory of S_n supplies us with explicit formulas for $2d$-juntas in $V_{d,\lambda}$ (arising in the construction of Specht modules), which can be turned into d-juntas by symmetrization. Since we will not need such explicit formulas here, we skip this description.

5.2 Finding a basis for $V_{A,B}$

We now move on to the study of the spaces $V_{A,B}$. These spaces have small dimension and are therefore easy to analyse. We first construct a set $\{v_T\}$ of functions in $V_{A,B}$ that form a nearly-orthonormal basis.

Definition 5.6. *Let $T = \{(i_1, j_1), \ldots, (i_k, j_k)\} \subseteq [d]^2$ be consistent. Let 1_T be the indicator function of permutation π in S_n that satisfy the restrictions given by T, i.e. $\pi(i_1) = j_1, \ldots, \pi(i_{i_k}) = j_k$. We define $v_T = \frac{1_T}{\|1_T\|_2}$.*

Since the spaces $V_{A,B}$ are isomorphic (as S_{n-d} bimodules) for all sets A, B of size d, we shall focus on the case where $A = B = [d]$.

Lemma 5.7. *Let $d \leqslant \frac{n}{2}$, and let $T \neq S$ be sets of size d. Then $\langle v_T, v_S\rangle \leqslant O\left(\frac{1}{n}\right)$.*

Proof. If $T \cup S$ is not consistent, then $1_T 1_S = 0$ and so $\langle v_T, v_S\rangle = 0$. Otherwise,

$$\langle v_T, v_S\rangle = \frac{\mathbb{E}\,|1_{T\cup S}|}{\|1_T\|_2\|1_S\|_2} = \frac{(n-|T\cup S|)!}{\sqrt{(n-|T|)!\,(n-|S|)!}} \leqslant \frac{(n-d-1)!}{(n-d)!} = O\left(\frac{1}{n}\right). \qquad \square$$

Proposition 5.8. *There exists an absolute constant $c > 0$ such that for all consistent $T \subseteq L$ we have*

$$\left\langle \mathrm{T}^{(\rho)} v_T, v_T \right\rangle \geqslant (c\rho)^{|T|}.$$

Proof. Let $x \sim L^m, y \sim N_\rho(x)$, and let $\sigma_x, \sigma_y \in S_n$ be corresponding permutations chosen according to the coupling. We have

$$\left\langle \mathrm{T}^{(\rho)} v_T, v_T \right\rangle = \frac{n!}{(n-|T|)!} \left\langle \mathrm{T}^{(\rho)} 1_T, 1_T \right\rangle,$$

as $\|1_T\|_2^2 = \frac{(n-|T|)!}{n!}$. We now interpret $\left\langle \mathrm{T}^{(\rho)} 1_T, 1_T \right\rangle$ as the probability that both σ_x and σ_y satisfy the restrictions given by T. For each ordered subset $S \subseteq [2n]$ of size $|T|$ consider the event A_S that $x_S = y_S = T$, while all the coordinates of the vectors $x_{[2n]\setminus S}, y_{[2n]\setminus S}$ do not contradict T and do not belong to T. Then

$$\left\langle \mathrm{T}^{(\rho)} x_T, x_T \right\rangle \geqslant \sum_{S \text{ an ordered } |T|\text{-subset of } [2n]} \Pr\left[A_S\right].$$

Now the probability that $x_S = T$ is $\left(\frac{1}{n}\right)^{2|T|}$. Conditioned on $x_S = T$, the probability that $y_S = T$ is at least $\rho^{|T|}$. When we condition on $x_S = y_S = T$, we obtain that the probability that $x_{[n]\setminus S}$ and $y_{[n]\setminus S}$ do not involve any coordinate contradicting T or in T is at least $\left(1 - \frac{2|T|}{n}\right)^{2n} = 2^{-\Theta(|T|)}$. Hence $\Pr\left[A_S\right] \geqslant \left(\frac{1}{n}\right)^{2|T|} \Omega\left(\rho\right)^{|T|}$. So wrapping everything up we obtain that

$$\left\langle \mathrm{T}^{(\rho)} v_T, v_T \right\rangle \geqslant \frac{(2n)!}{(2n-|T|)!} \cdot \frac{n!}{(n-|T|)!} \frac{1}{n^{2|T|}} \Omega(\rho)^{|T|} = \Omega\left(\rho\right)^{|T|}. \qquad \square$$

Lemma 5.9. *Let $\rho \in (0,1)$. Then for all sets $T \neq S$ of size at most $n/2$ we have $\left\langle \mathrm{T}^{(\rho)} v_T, v_S \right\rangle = O\left(\frac{1}{\sqrt{n}}\right)$.*

Proof. Suppose without loss of generality that $\|1_T\|_2^2 \leqslant \|1_S\|_2^2$, so $|T| \geqslant |S|$. Choose $x \sim L^m, y \sim N_\rho(x)$, and let σ_x, σ_y by the corresponding random permutations given by the coupling. We have

$$\left\langle \mathrm{T}^{(\rho)} v_T, v_S \right\rangle = \frac{\Pr\left[1_T\left(\sigma_x\right) = 1, 1_S\left(\sigma_y\right) = 1\right]}{\sqrt{\mathbb{E} 1_T \mathbb{E} 1_S}}$$

As the probability in the numerator is at most $\mathbb{E}\left[1_T\right]$, we have

$$\left\langle T^{(\rho)} v_T, v_S \right\rangle \leqslant \sqrt{\frac{\mathbb{E}\left[1_T\right]}{\mathbb{E}\left[1_S\right]}} = \sqrt{\frac{(n-|T|)!}{(n-|S|)!}},$$

and the proposition follows in the case that $|S| < |T|$.

It remains to prove the proposition provided that $|S| = |T|$. Let $(i,j) \in S \setminus T$. Note that

$$\Pr\left[1_T\left(\sigma_x\right) = 1, 1_S\left(\sigma_y\right) = 1\right] \leqslant \frac{1}{n} \Pr\left[1_T\left(\sigma_x\right) = 1 \mid \sigma_y(i) = j\right].$$

Let us condition further on $\sigma_x(i)$. Conditioned on $\sigma_x(i) = j$, we have that σ_x is a random permutation sending i to j, and so $\Pr\left[1_T\left(\sigma_X\right) = 1\right]$ is either 0 (if (i,j) contradicts T) or $\frac{(n-1-|T|)!}{(n-1)!} = O\left(\|1_T\|_2^2\right)$ (if (i,j) is consistent with T).

Conditioned on $\sigma_x(i) \neq j$ (and on $\sigma_y(i) = j$), we again obtain that σ_x is a random permutation that does not send i to j, in which case

$$\Pr\left[1_T\left(\sigma_x\right) = 1\right] = \frac{(n-|T|)!}{n! - (n-1)!} = O\left(\|1_T\|_2^2\right)$$

if (i,j) contradicts T, and

$$\Pr\left[1_T\left(\sigma_x\right)=1\right]=\frac{(n-|T|)!-(n-|T|-1)!}{n!-(n-1)!}=O\left(\|1_T\|_2^2\right)$$

if (i,j) is consistent with T. This completes the proof of the lemma. □

Proposition 5.10. *Let C be a sufficiently large constant. If $n\geqslant\left(\frac{\rho}{C}\right)^{-d}C^{d^2}$ and f is a d-junta, then*

$$\left\langle \mathrm{T}^{(\rho)}f,f\right\rangle\geqslant\rho^{O(d)}\|f\|_2^2.$$

Proof. Since $\{v_T\}_{T\subseteq[d]^2}$ span the space $V_{[d],[d]}$ of $([d],[d])$-juntas by Lemma 2.3, we may write $f=\sum a_T v_T$. Now

$$\left\langle \mathrm{T}^{(\rho)}f,f\right\rangle=\sum_T a_T^2\left\langle \mathrm{T}^{(\rho)}v_T,v_T\right\rangle+\sum_{T\neq S}a_Ta_S\left\langle \mathrm{T}^{(\rho)}v_T,v_S\right\rangle.$$

By Lemma 5.9 we have

$$\left|\sum_{T\neq S}a_Ta_S\left\langle \mathrm{T}^{(\rho)}v_T,v_S\right\rangle\right|\leqslant O\left(\sum_{T\neq S}\frac{|a_Ta_S|}{\sqrt{n}}\right)\leqslant O\left(\frac{1}{\sqrt{n}}\right)\left(\sum_T|a_T|\right)^2\leqslant\frac{2^{O\left(d^2\right)}}{\sqrt{n}}\left(\sum_T|a_T|^2\right),$$

where the last inequality is by Cauchy–Schwarz. On the other hand, by Proposition 5.8 we have

$$\sum_T a_T^2\left\langle \mathrm{T}^{(\rho)}v_T,v_T\right\rangle\geqslant\rho^{O(d)}\left(\sum_T a_T^2\right).$$

Using a similar calculation, one sees that

$$\|f\|_2^2=\left(1\pm\frac{2^{O\left(d^2\right)}}{n}\right)\sum_T a_T^2,$$

so we get that

$$\left\langle \mathrm{T}^{(\rho)}f,f\right\rangle\geqslant\left(\rho^{O(d)}-\frac{2^{O\left(d^2\right)}}{\sqrt{n}}\right)\sum_T a_T^2\geqslant\left(\rho^{O(d)}-\frac{2^{O\left(d^2\right)}}{\sqrt{n}}\right)\|f\|_2^2\geqslant\rho^{O(d)}\|f\|_2^2.$$ □

Corollary 5.11. *Let C be a sufficiently large absolute constant. If $n\geqslant\left(\frac{\rho}{C}\right)^{-d}C^{d^2}$ then all the eigenvalues of $\mathrm{T}^{(\rho)}$ as an operator from V_d to itself are at least $\rho^{O(d)}$.*

Proof. By Lemma 5.5, each eigenspace $V_{d,\lambda}$ contains a d-junta. Let $f\in V_{d,\lambda}$ be a nonzero d-junta. Then by Proposition 5.10,

$$\lambda=\frac{\left\langle \mathrm{T}^{(\rho)}f,f\right\rangle}{\|f\|_2^2}\geqslant\rho^{O(d)}.$$ □

5.3 Showing that the eigenvalues of $\mathrm{T}^{(\rho)}$ on V_d are concentrated on at most d values

Let $\lambda_i(\rho) = \left\langle \mathrm{T}^{(\rho)} v_T, v_T \right\rangle$, where T is a set of size i. Then symmetry implies that $\lambda_i(\rho)$ does not depend on the choice of T.

Lemma 5.12. *Suppose that $n \geqslant \left(\frac{\rho}{C}\right)^{O(d)} C^{d^2}$. Then each eigenvalue of $\mathrm{T}^{(\rho)}$ as an operator on V_d is equal to $\lambda_i(\rho)\left(1 \pm n^{-\frac{1}{3}}\right)$ for some $i \leqslant d$.*

Proof. Let λ be an eigenvalue of $\mathrm{T}^{(\rho)}$, 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 $0 = \mathrm{T}^{(\rho)} f - \lambda f$, but on the other hand for each set S we have

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

Thus, for all S we have that

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

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

5.4 An L^2 variant of Lemma 3.6

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

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

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

Suppose the maximum is attained at $\lambda_\star$. By Lemma 5.12, there is $i \leqslant d$ such that $\lambda_\star = \lambda_i(1 \pm n^{-\frac{1}{3}})$, and so

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

For any $j \neq i$, we have by Proposition 5.10 that $\lambda_j \geqslant \rho^{O(d)}$, and so

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

Combining the two inequalities, we get that

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

where the last inequality follows from the lower bound on n. To finish up the proof then, we must upper bound $\|P\|$, and this is relatively straightforward:

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

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

5.5 Deducing the L^q approximation

To deduce the L^q approximation of the polynomial P from Lemma 5.13 we use the following basic type of hypercontractive inequality (this bound is often times too weak quantitatively, but it is good enough for us since we have a very strong L_2 approximation).

Lemma 5.14. *Let C be sufficiently large, and let $n \geqslant C^{d^2} q^{2d}$. Let $f\colon S_n \to \mathbb{R}$ be a function of degree d. Then $\|f\|_q \leqslant q^{O(d)} n^d \|f\|_2$.*

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 5.11, we get that $\|g\|_2 \leqslant \rho^{-O(d)} \|f\|_2$. Thus, we have that $\|f\|_q = \left\|T^{(\rho)} g\right\|_q$, and to upper bound this norm we intend to use Theorem 3.3, and for that we show that g is global with fairly weak parameters.

Let $T \subseteq L$ be consistent of size at most $2d$. Then

$$\|g_{\to T}\|_2^2 = \frac{\mathbb{E}_x\, g(x) 1_T(x)}{\mathbb{E}_x[1_T(x)]} \leqslant \sqrt{\frac{\mathbb{E}_x\, g(x)^2}{\mathbb{E}_x[1_T(x)]}} \leqslant n^{\frac{|T|}{2}} \|g\|_2^2 \leqslant n^{\frac{|T|}{2}} \rho^{-O(d)} \|f\|_2^2,$$

and so g is $(2d, \varepsilon)$ global for $\varepsilon = n^{d/2} \rho^{-O(d)} \|f\|_2$. Lemma 3.5 now implies that g is ε-global with constant 4^8. By the choice of ρ, we may now use Theorem 3.3 to deduce that

$$\left\|T^{(\rho)} g\right\|_q \leqslant \varepsilon^{(q-2)/q} \|g\|_2^{2/q} \leqslant n^{d/2} \rho^{-O(d)} \|f\|_2 \leqslant n^d q^{O(d)} \|f\|_2 . \qquad \square$$

Finally, we combine Lemma 5.13 and Lemma 5.14 to deduce the L^q approximating polynomial.

Proof of Lemma 3.6. Let f be a function of degree d. By lemma 5.13 there exists a P with $\|P\| \leqslant \rho^{-O\left(d^3\right)}$ and $P(0) = 0$ such that the function $g = P\left(\mathrm{T}^{(\rho)}\right) f - f$ satisfies $\|g\|_2 \leqslant n^{-2d} \|f\|_2$. By Lemma 5.14, $\|g\|_q \leqslant q^{4d} n^{-d} \|f\|_2 \leqslant \frac{1}{\sqrt{n}} \|f\|_2$, provided that C is sufficiently large, completing the proof. □

6 Hypercontractivity: the direct approach

In this section, we give an alternative proof to a variant of Theorem 1.4. This approach starts by identifying a trivial spanning set of the space V_t of degree t functions from Definition 2.1.

Notations. For technical reasons, it will be convenient for us to work with ordered sets. We denote by $[n]_t$ the collection of ordered sets of size t, which are simply t-tuples of distinct elements from $[n]$, but we also allow set operations (such as $\setminus$) on them. We also denote $n_t = |[n]_t| = n(n-1)\cdots(n-t+1)$. For ordered sets $I = \{i_1, \dots, i_t\}$, $J = \{j_1, \dots, j_t\}$, we denote by $1_{I\to J}(\pi)$ the indicator of $\pi(i_k) = j_k$ for all $k = 1, \dots, t$; for convenience, we also denote this by $\pi(I) = J$.

With the above notations, the following set clearly spans V_t, by definition:

$$\{1_{I\to J} \mid |I| = |J| \leqslant t\}. \tag{4}$$

We remark that this set is not a basis, since these functions are linearly dependent. For example, for $t = 1$ we have $\sum_{i=1}^{n} 1_{\pi(1)=i} - 1 = 0$. This implies that a function $f \in V_1$ has several different representations as a linear combination of functions from the spanning set (4). The key to our approach is to show that there is a way to canonically choose such a linear combination, which is both unique and works well with computations of high moments.

Definition 6.1. *Let $f \in V_{=t}$, and suppose that $f = \sum\limits_{I,J\in[n]_t} a(I,J)1_{I\to J}$. We say that this representation is* normalized *if*

1. *For any $1 \leqslant r \leqslant t$, $J = \{j_1, \dots, j_t\}$ and $I = \{i_1, \dots, i_{r-1}, i_{r+1}, \dots, i_t\}$ we have that*

$$\sum_{i_r \notin I} a(\{i_1, \dots, i_t\}, J) = 0.$$

2. *Analogously, for any $1 \leqslant r \leqslant t$, $I = \{i_1, \dots, i_t\}$ and $J = \{j_1, \dots, j_{r-1}, j_{r+1}, \dots, j_t\}$ we have that*

$$\sum_{j_r \notin J} a(I, \{j_1, \dots, j_t\}) = 0.$$

3. *Symmetry: for all ordered sets I, J of size t and $\pi \in S_t$, we have $a(I,J) = a(\pi(I), \pi(J))$.*

More loosely, we say that a representation according to the spanning set (4) is normalized if averaging the coefficients according to a single coordinate results in 0. We also refer to the equalities in Definition (4) as "normalizing relations". In this section, we show that a normalized representation always exists, and then show how it is useful in establishing hypercontractive statements similar to Theorem 1.4.

Normalized representations first appear in the context of the slice by Dunkl [5], who called normalized representations *harmonic functions*. See also the monograph of Bannai and Ito [1, III.3] and the papers [8, 9]. Ryan O'Donnell (personal communication) has proposed calling them *zero-flux* representations.

6.1 Finding a normalized representation

Lemma 6.2. *Let $0 \leqslant t \leqslant t$, and let $f \in V_t$. Then we may write $f = h + g$, where $h \in V_{t-1}$ and g is given by a set of coefficients satisfying the normalizing relations $g = \sum\limits_{I,J\in[n]_t} a_t(I,J)1_{I\to J}(\pi)$.*

Proof. The proof is by induction on t.

Fix $t \geqslant 1$ and $f \in V_t$. Then we may write $f(\pi) = \sum\limits_{I,J\in[n]_t} a(I,J)1_{I\to J}(\pi)$, where the coefficients satisfy the symmetry property from Definition 6.1.

Throughout the proof, we will change the coefficients in a sequential process, and always maintain the form $f = h + \sum_{|I|=|J|=t} b(I,J)1_{I\to J}(\pi)$ for $h \in V_{t-1}$.

Take $r \in [t]$, and for each $I = \{i_1, \ldots, i_t\}$, $J = \{j_1, \ldots, j_t\}$, define the coefficients

$$b(I,J) = a(I,J) - \frac{1}{n-t+1} \sum_{i \notin I \setminus \{i_r\}} a(\{i_1, \ldots, i_{r-1}, i, i_{r+1}, \ldots, i_t\}, J). \tag{5}$$

In Claim 6.3 below, we prove that after making this change of coefficients, we may write $f = h + \sum_{|I|=|J|=t} b(I,J)1_{I\to J}(\pi)$, and that the coefficients $b(I,J)$ satisfy all normalizing relations that the $a(I,J)$ do, as well as the normalizing relations from the first collection in Definition 6.1 for r. We repeat this process for all $r \in [t]$.

After this process is done, we have $f = h + \sum_{I,J\in[n]_t} b(I,J)1_{I\to J}(\pi)$, where the coefficients $a(I,J)$ satisfy the first collection of normalizing relations from Definition 6.1. We can now perform the analogous process on the J part, and by symmetry obtain that after this process, the second collection of normalizing relations in Definition 6.1 hold. One only has to check that this does not destroy the first collection of normalizing relations, which we also prove in Claim 6.3.

Finally, we symmetrize f to ensure that it satisfies the symmetry condition. To do so, we replace $g = \sum_{I,J\in[n]_t} b(I,J)1_{I\to J}(\pi)$ with $g' = \sum_{\pi\in S_t} g^{\pi}$, where $(1_{I\to J})^{\pi} = 1_{\pi(I)\mapsto\pi(J)}$ (and extended linearly). It is easy to check that $g = g^{\pi}$ as functions, and that g^{π} satisfies the two sets of normalizing relations. It follows that so does g', and furthermore by construction, g' is symmetric. $\square$

Claim 6.3. *The change of coefficients* (5) *has the following properties:*

1. *The coefficients $b(I,J)$ satisfy the normalizing relation in the first item for r in Definition 6.1.*
2. *If the coefficients $a(I,J)$ satisfy the normalizing relation in the first item in Definition 6.1 for $r' \neq r$, then so do $b(I,J)$.*
3. *If the coefficients $a(I,J)$ satisfy the normalizing relation in the second item in Definition 6.1 for r', then so do $b(I,J)$.*
4. *We may write $f = h + \sum_{|I|=|J|=t} b(I,J)1_{I\to J}(\pi)$, where $h \in V_{t-1}$.*

Proof. We prove each one of the items separately.

Proof of the first item. Fix $I = \{i_1, \ldots, i_{r-1}, i_{r+1} \ldots, i_t\}$, $J = \{j_1, \ldots, j_t\}$, and calculate:

$$\begin{aligned}\sum_{i_r\notin I} b(\{i_1,\ldots,i_t\},J) &= \sum_{i_r\notin I}\left(a(\{i_1,\ldots,i_t\},J) - \frac{1}{n-t+1}\sum_{i\notin I} a(\{i_1,\ldots,i_{r-1},i,i_{r+1},\ldots,i_t\},J)\right)\\ &= \sum_{i_r\notin I} a(\{i_1,\ldots,i_t\},J) - \frac{1}{n-t+1}\sum_{\substack{i_r\notin I\\ i\notin I}} a(\{i_1,\ldots,i_{r-1},i,i_{r+1},\ldots,i_t\},J).\end{aligned} \tag{6}$$

As in the second double sum, for each i_r the coefficient $a(\{i_1, \ldots, i_{r-1}, i_r, i_{r+1}, \ldots, i_t\}, J)$ is counted $n - |I| = n - t + 1$ times, we get that the above expression is equal to 0.

Proof of the second item. Fix $r' \neq r$, and suppose $a(\cdot,\cdot)$ satisfy the first set of normalizing relations for r'. Without loss of generality, assume $r' < r$. Let $I = \{i_1, \ldots, i_{r'-1}, i_{r'+1}, \ldots, i_t\}$, $J = \{j_1, \ldots, j_t\}$. Below, we let $i, i_{r'}$ be summation indices and we denote $I' = \{i_1, \ldots, i_{r'-1}, i_{r'}, i_{r'+1}, \ldots, i_{r-1}, i, i_r, \ldots, , i_t\}$. Calculating as in (6):

$$\begin{aligned}\sum_{i_{r'} \notin I} b(\{i_1, \ldots, i_t\}, J) &= \sum_{i_{r'} \notin I} a(\{i_1, \ldots, i_t\}, J) - \frac{1}{n-t+1} \sum_{i \notin I \setminus \{i_r\}} a(I', J) \\ &= \sum_{i_{r'} \notin I} a(\{i_1, \ldots, i_t\}, J) - \frac{1}{n-t+1} \sum_{i_{r'} \notin I} \sum_{i \notin I \setminus \{i_r\}} a(I', J). \end{aligned} \tag{7}$$

The first sum is 0 by the assumption of the second item. For the second sum, we interchange the order of summation to see that it is equal to $\sum_{i \notin I \setminus \{i_r\}} \sum_{i_{r'} \notin I} a(I', J)$, and note that for each i, the inner sum is 0 again by the assumption of the second item.

Proof of the third item. Fix r', and suppose $a(\cdot,\cdot)$ satisfy the second set of normalizing relations for r'. Fix $I = \{i_1, \ldots, i_t\}$, $J = \{j_1, \ldots, j_{r'-1}, j_{r'+1}, \ldots, j_t\}$, $I' = \{i_1, \ldots, i_{r-1}, i, i_{r+1}, \ldots, i_t\}$, $J' = \{j_1, \ldots, j_t\}$, and calculate:

$$\begin{aligned}\sum_{j_r \notin J} b(I, J') &= \sum_{j_r \notin J} \left(a(I, J') - \frac{1}{n-t+1} \sum_{i \notin I \setminus \{i_r\}} a(I', J') \right) \\ &= \sum_{j_r \notin J} a(I, J') - \frac{1}{n-t+1} \sum_{i \notin I \setminus \{i_r\}} \sum_{j_r \notin J} a(I', J'). \end{aligned} \tag{8}$$

Once again, both sums vanish due to the assumption.

Proof of the fourth item. For $I = \{i_1, \ldots, i_t\}$, $J = \{j_1, \ldots, j_t\}$, denote

$$c(I, J) = \frac{1}{n-t+1} \sum_{i \notin I \setminus \{i_r\}} a(\{i_1, \ldots, i_{r-1}, i, i_{r+1}, \ldots, i_t\}, J),$$

so that $a(I, J) = b(I, J) + c(I, J)$. Plugging this into the representation of f, we see that it is enough to prove that $h(\pi) = \sum_{I,J} c(I, J) 1_{I \to J}(\pi)$ is in V_{t-1}. Writing $I' = I \setminus \{i_r\}$, $J' = J \setminus \{j_r\}$ and expanding, we see that

$$\begin{aligned} h(\pi) &= \frac{1}{n-t+1} \sum_{I,J} 1_{I \to J}(\pi) \sum_{i \notin I \setminus \{i_r\}} a(\{i_1, \ldots, i_{r-1}, i, i_{r+1}, \ldots, i_t\}, J) \\ &= \frac{1}{n-t+1} \sum_{I',J'} \sum_{i \notin I', j_r \notin J'} a(\{i_1, \ldots, i_{r-1}, i, i_{r+1}, \ldots, i_t\}, J) \sum_{i_r \notin I'} 1_{I \to J}(\pi). \end{aligned}$$

Noting that $\sum_{i_r \notin I'} 1_{I \to J}(\pi) = 1_{I' \to J'}(\pi)$ is in the spanning set (4) for $t-1$, the proof is concluded. □

Applying Lemma 6.2 iteratively, we may write each $f \colon S_n \to \mathbb{R}$ of degree at most t as $f = f_0 + \ldots + f_d$, where for each $k = 0, 1, \ldots, d$, the function f_k is in V_k, and is given by a list of coefficients satisfying the normalizing relations.

6.2 Usefulness of normalized representations

In this section we establish a claim that demonstrates the usefulness of the normalizing relations. Informally, this claim often serves as a replacement for the orthogonality property that is so useful in product spaces. Formally, it allows us to turn long sums into short sums, and is very helpful in various computations arising in computations in norms of functions on S_n that are given in a normalized representation.

Claim 6.4. *Let $r \in \{1, \ldots, d\}$, $0 \leqslant t < r$. Let J be of size r, I be of size at least r, and $R \subseteq I$ of size $r - t$. Then*

$$\sum_{T \in ([n] \setminus I)_t} a_r(R \circ T, J) = (-1)^t \sum_{T \in (I \setminus R)_t} a_r(R \circ T, J).$$

Proof. By symmetry, it suffices to prove the statement for R that are prefixes of I. We prove the claim by induction on t. The case $t = 0$ is trivial, so assume the claim holds for $t - 1$, where $t \geqslant 1$, and prove for t. The left hand side is equal to

$$\sum_{\substack{i_1, \ldots, i_t \notin I \\ \text{distinct}}} a_r(R \circ (i_1, \ldots, i_t), J).$$

For fixed $i_1, \ldots, i_{t-1} \notin I$, by the normalizing relations we have that

$$\sum_{i_t \notin I \cup \{i_1, \ldots, i_{t-1}\}} a_r(R \circ (i_1, \ldots, i_{t-1}) \circ (i_t), J) = -\sum_{i_t \in I \setminus R} a_r(R \circ (i_1, i_2, \ldots, i_{t-1}) \circ (i_t), J),$$

hence

$$\sum_{\substack{i_1, \ldots, i_t \notin I \\ \text{distinct}}} a_r(R \circ (i_1, \ldots, i_t), J) = -\sum_{i_t \in I \setminus R} \sum_{\substack{i_1, \ldots, i_{t-1} \notin I \cup \{i_t\} \\ \text{distinct}}} a_r(R \circ (i_1, i_2, \ldots, i_{t-1}) \circ i_t, J).$$

For fixed $i_t \in I \setminus R$, using the induction hypothesis, the inner sum is equal to

$$(-1)^{t-1} \sum_{T \in (I \setminus (R \cup \{i_t\}))_{t-1}} a_r(R \circ T \circ (i_t), J).$$

Plugging that in,

$$\begin{aligned} \sum_{\substack{i_1, \ldots, i_t \notin I \\ \text{distinct}}} a_r(R \circ (i_1, \ldots, i_t), J) &= (-1)^{t-1} \sum_{i_t \in I \setminus R} - \sum_{T \in (I \setminus (R \cup \{i_t\}))_{t-1}} a_r(R \circ T \circ (i_t), J) \\ &= (-1)^t \sum_{T' \in (I \setminus R)_t} a_r(R \circ T', J). \end{aligned}$$

$\square$

6.3 Analytic influences and the hypercontractive statement

Key to the hypercontractive statement proved in this section is an analytic notion of influence. Given a fixed representation of f as $\sum\limits_{k=0}^{n} \sum\limits_{I, J \in [n]_k} a_k(I, J) 1_{I \to J}$ where for each k the coefficients $a_k(I, J)$ satisfy the normalizing relations, we define the analytic notion of influences as follows.

Definition 6.5. *For $S, T \subseteq [n]$ of the same size s, define*

$$I_{S,T}[f] = \sum_{r \geqslant 0} \sum_{\substack{I \in ([n] \setminus S)_r \\ J \in ([n] \setminus T)_r}} (r+s)!^2 \frac{1}{n^{r+s}} a(S \circ I, T \circ J)^2.$$

Here, $S \circ I$ denotes the element in $[n]_r$ resulting from appending I at the end of S.

Definition 6.6. *A function f is called ε-analytically-global if for all S, T, $I_{S,T}[f] \leqslant \varepsilon$.*

Remark 6.7. *With some work it can be shown that for $d \ll n$, a degree d function being ε-analytically global is equivalent to f being $(2d, \delta)$-global in the sense of Definition 1.3, where $\delta = O_d(\varepsilon)$. Thus, at least qualitatively, the hypercontractive statement below is in fact equivalent to Theorem 1.4.*

We can now state our variant of the hypercontractive inequality that uses analytic influences.

Theorem 6.8. *There exists an absolute constant $C > 0$ such that for all $d, n \in \mathbb{N}$ for which $n \geqslant 2^{C \cdot d \log d}$, the following holds. If $f \in V_d$ is given by a list of coefficients satisfying the normalizing relations, say $f = \sum\limits_{I,J \in [n]_d} a_d(I,J) 1_{I \to J}$, then*

$$\mathbb{E}_{\pi}\left[f(\pi)^4\right] \leqslant \sum_{|S|=|T|} \left(\frac{4}{n}\right)^{|S|} I_{S,T}[f]^2.$$

p-biased hypercontractivity. The last ingredient we use in our proof is a hypercontractive inequality on the p-biased cube from [17]. Let $g\colon \{0,1\}^m \to \mathbb{R}$ be a degree d function, where we think of $\{0,1\}^m$ as equipped with the p-biased product measure. Then, we may write g in the basis of characters, i.e. as a liner combination of $\{\chi_S\}_{S \subseteq [m]}$, where $\chi_S(x) = \prod\limits_{i \in S} \frac{x_i - p}{\sqrt{p(1-p)}}$. This is the p-biased Fourier transform of f:

$$g(x) = \sum_{S} \widehat{g}(S) \chi_S(x).$$

Next, we define the generalized influences of sets (which are very close in spirit to the analytic notion of influences considered herein). For $T \subseteq [n]$, we denote

$$I_T[g] = \sum_{S \supseteq T} \widehat{g}(S)^2.$$

The following results is an easy consequence of [17, Theorem 3.4] (the deduction of it from this result is done in the same way as the proof of [17, Lemma 3.6]).

Theorem 6.9. *Suppose $g\colon \{0,1\}^m \to \mathbb{R}$. Then $\|g\|_4^4 \leqslant \sum\limits_{T \subseteq [n]} (3p)^{|T|} I_T[g]^2$.*

6.4 Proof of Theorem 6.8

Write f according to its normalized representation as $f(\pi) = \sum\limits_{I,J \in [n]_d} a(I,J) 1_{I \to J}$. We intend to define a function $g\colon \{0,1\}^{n \times n} \to \mathbb{R}$ that will behaves similary to f, as follows. We think of $\{0,1\}^{n \times n}$ as equipped with the p-biased measure for $p = 1/n$, and think of an input $x \in \{0,1\}^{n\times n}$ as a matrix. The rationale is that the bit $x_{i,j}$ being 1 will encode the fact that $\pi(i) = j$, but we will never actually think about it this way. Thus, we define g as

$$g(x) = \sum_{I,J\in[n]_d} a(I,J) \prod_{\ell=1}^{d} \left(1_{I_\ell \to J_\ell} - \frac{1}{n}\right).$$

For I, J, we denote by $S_{I,J} \subseteq [n\times n]$ the set of coordinates $\{\,(I_\ell, J_\ell) \mid \ell = 1,\dots,d\}$, and note that with this notation,

$$g(x) = \sum_{I,J\in[n]_d} \sqrt{p(1-p)}^{d}\, |a(I,J)|\, \chi_{S_{I,J}}(x).$$

To complete the proof, we first show (Claim 6.10) that $\|f\|_4^4 \leqslant (1+o(1))\,\|g\|_4^4$, and then prove the desired upper bound on the 4-norm of g, using Theorem 6.9.

Claim 6.10. $\|f\|_4^4 \leqslant (1+o(1))\,\|g\|_4^4$

Proof. Deferred to Section 6.4.1. □

We now upper bound $\|g\|_4^4$. Using Theorem 6.9,

$$\|g\|_4^4 \leqslant \sum_{T\subseteq[n\times n]} (3p)^{|T|} I_T[g]^2, \tag{9}$$

and the next claim bounds the generalized influences of g by the analytic influences of f.

For two sets $I = \{i_1,\dots,i_t\}$, $J = \{j_1,\dots,j_t\}$ of the same size, let $S(I,J) = \{(i_1,j_1),\dots,(i_t,j_t)\} \subseteq [n]\times[n]$.

Claim 6.11. *Let $T = S(I',J')$ be such that $I_T[g] \neq 0$. Then $I_T[g] \leqslant I_{I',J'}[f]$.*

Proof. Take T in this sum for which $I_T[g] \neq 0$, and denote $t = |T|$. Then $T = \{(i_1,j_1),\dots,(i_t,j_t)\} = S(I',J')$ for $I' = \{i_1,\dots,i_t\}$, $J' = \{j_1,\dots,j_t\}$ that are consistent. For $Q \subseteq [n]\times[n]$ of size d such that $T \subseteq Q$, let $S_{Q,T} = \{\,(I,J) \mid T \subseteq S(I,J) = Q\}$, and note that by the symmetry normalizing relation, $a(I,J)$ is constant on $(I,J) \in S_{Q,T}$. We thus get

$$I_T[g] = \sum_{Q} \left(\sum_{(I,J)\in S_{Q,T}} \sqrt{p(1-p)}^{d} a(I,J)\right)^2 \leqslant d!p^d \sum_{Q} \sum_{(I,J)\in S_{Q,T}} a(I,J)^2,$$

where we used the fact that the size of $S_{Q,T}$ is $d!$. Rewriting the sum by first choosing the locations of T in (I,J), we get that the last sum is at most

$$d_t \sum_{\substack{I\in([n]\setminus I')_{d-t}\\ J\in([n]\setminus J')_{d-t}}} a(I'\circ I, J'\circ J)^2$$

Combining all, we get that $I_T[g] \leqslant \sum_{\substack{I\in([n]\setminus I')_{d-t}\\ J\in([n]\setminus J')_{d-t}}} d!^2 \frac{1}{n^d} a(I'\circ I, J'\circ J)^2 = I_{I',J'}[g]$. □

Plugging in Claim 6.11 into (9) and using Claim 6.10 finishes the proof of Theorem 6.8.

6.4.1 Proof of Claim 6.10

Let I_r and J_r be d-tuples of distinct indices from $[n]$. Then

$$\mathop{\mathbb{E}}_{\pi}\left[f(\pi)^4\right] = \sum_{\substack{I_1,\dots,I_4\\J_1,\dots,J_4}} a(I_1,J_1)\cdots a(I_4,J_4)\mathop{\mathbb{E}}_{\pi}\left[1_{\pi(I_1)=J_1}\cdots 1_{\pi(I_4)=J_4}\right].$$

Consider the collection of constraints on π in the product of the indicators. To be non-zero, the constraints should be consistent, so we only consider such tuples. Let M be the number of different elements that appear in $I_1,\dots,I_4$ (which is at least d and at most $4d$) We partition the outer sum according to M, and upper bound the contribution from each M separately. Fix M; then the contribution from it is:

$$\frac{1}{n_M}\sum_{\substack{I_1,\dots,I_4\\J_1,\dots,J_4\\\text{type } M}} a(I_1,J_1)\cdots a(I_4,J_4).$$

We would like to further partition this sum according to the pattern in which the M different elements of $I_1,\dots,I_4$ are divided between them (and by consistency, this determines the way the M different elements of $J_1,\dots,J_4$ are divided between them). There are at most $(2^4-1)^M \leqslant 2^{16d}$ different such configurations, thus we fix one such configuration and upper bound it (at the end multiplying the bound by 2^{16d}). Thus, we have distinct $i_1,\dots,i_M$ ranging over $[n]$, and the coordinate of each I_r is composed of the $i_1,\dots,i_M$ (and similarly $j_1,\dots,j_M$ and the J_r's), and our sum is

$$\frac{1}{n_M}\sum_{\substack{i_1,\dots,i_M \text{ distinct}\\j_1,\dots,j_M \text{ distinct}}} a(I_1,J_1)\cdots a(I_4,J_4). \tag{10}$$

We partition the i_t's into the number of times they occur: let $A_1,\dots,A_4$ be the sets of i_t that appear in $1,2,3,$ or 4 of the I_r's. We note that i_t and j_t appear in the same I_r's and always together (otherwise the constraints would be contradictory), and in particular $i_t \in A_j$ iff $j_t \in A_j$. Also, $M = |A_1|+|A_2|+|A_3|+|A_4|$.

We consider contributions from configurations where $A_1=\emptyset$ and $A_1\neq\emptyset$ separately, and to control the latter group we show that the above sum may be upper bounded by M^{2M} sums of in which $A_1=\emptyset$. To do that, we show how to reduce the size of A_1 by allowing more sums, and then apply it iteratively.

Without loss of generality, assume $i_1 \in A_1$; then it is in exactly one of the I_r's — without loss of generality the last coordinate of I_4. We rewrite the sum as

$$\frac{1}{n_M}\sum_{i_1,\dots,i_M} a(I_1,J_1)a(I_2,J_2)a(I_3,J_3)\sum_{\substack{i_1\in[n]\setminus\{i_2,\dots,i_M\}\\j_1\in[n]\setminus\{j_2,\dots,j_M\}}} a(I_4,J_4). \tag{11}$$

Consider the innermost sum. Applying Claim 6.4 twice, we have

$$\sum_{\substack{i_1\in[n]\setminus\{i_2,\dots,i_M\}\\j_1\in[n]\setminus\{j_2,\dots,j_M\}}} a(I_4,J_4) = \sum_{\substack{i_1\in\{i_2,\dots,i_M\}\setminus I_4\\j_1\in\{j_2,\dots,j_M\}\setminus J_4}} a(I_4,J_4).$$

Plugging that into (11), we are able to write the sum therein using $(M-r)^2$ sums (one for each choice of $i_1\in\{i_2,\dots,i_M\}\setminus I_4$ and $j_1\in\{j_2,\dots,j_M\}\setminus J_4$) on $i_2,\dots,i_M,j_2,\dots,j_M$, and thus we have reduced the size of A_1 by at least 1, and have decreased M by at least 1. The last bit implies that the original normalizing factor is smaller by a factor of at least $1/n$ than the new one. Iteratively applying this procedure, we end up with $A_1 = \emptyset$, and we assume that henceforth. Thus, letting $\mathcal{H}$ be the set of consistent $(I_1, \ldots, I_4, J_1, \ldots, J_4)$ in which each element in $I_1 \cup \cdots \cup I_4$ appears in at least two of the I_i's, we get that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\pi}\left[f(\pi)^4\right] &\leqslant \left(1 + \frac{d^{O(d)}}{n}\right) \sum_{\substack{I_1,\ldots,I_4 \\ J_1,\ldots,J_4 \\ \text{from } \mathcal{H}}} |a(I_1, J_1)| \cdots |a(I_4, J_4)| \mathop{\mathbb{E}}_{\pi}\left[1_{\pi(I_1)=J_1} \cdots 1_{\pi(I_4)=J_4}\right] \\
&\leqslant (1 + o(1)) \sum_{\substack{I_1,\ldots,I_4 \\ J_1,\ldots,J_4 \\ \text{from } \mathcal{H}}} \frac{1}{n^{|I_1 \cup \cdots \cup I_4|}} |a(I_1, J_1)| \cdots |a(I_4, J_4)|, \qquad (12)
\end{aligned}$$

where in the last inequality we used

$$\mathop{\mathbb{E}}_{\pi}\left[1_{\pi(I_1)=J_1} \cdots 1_{\pi(I_4)=J_4}\right] = \frac{1}{n \cdot (n-1) \cdots (n - |I_1 \cup \ldots \cup I_4| + 1)} \leqslant (1 + o(1)) \frac{1}{n^{|I_1 \cup \ldots \cup I_4|}}.$$

Next, we lower bound $\|g\|_4^4$. Expanding as before,

$$\mathop{\mathbb{E}}_{x}\left[g(\pi)^4\right] = \sum_{\substack{I_1,\ldots,I_4 \\ J_1,\ldots,J_4}} \sqrt{p(1-p)}^{4d} |a(I_1, J_1)| \cdots |a(I_4, J_4)| \mathop{\mathbb{E}}_{x}\left[\chi_{S(I_1,J_1)}(x) \cdots \chi_{S(I_4,J_4)}(x)\right].$$

A direct computation shows that the expectation of a normalized p-biased bit, i.e. $\frac{x_{i,j} - p}{\sqrt{p(1-p)}}$, is 0, the expectation of its square is 1, the expectation of its third power is $\frac{1+o(1)}{\sqrt{p(1-p)}}$, and the expectation of its fourth power is $\frac{1+o(1)}{p(1-p)}$. This tells us that all summands in the above formula are non-negative, and therefore we can omit all those that correspond to $(I_1, \ldots, I_4)$ and $(J_1, \ldots, J_4)$ not from $\mathcal{H}$, and only decrease the quantity. For $j = 2, 3, 4$, denote by h_j the number of elements that appear in j of the $I_1, \ldots, I_4$. Then we get that the inner term is at least

$$(1 - o(1)) \sqrt{p(1-p)}^{4d - h_3 - 2h_4} |a(I_1, J_1)| \cdots |a(I_4, J_4)|.$$

Note that $2h_2 + 3h_3 + 4h_4 = 4d$, we get that $4d - h_3 - 2h_4 = 2(h_2 + h_3 + h_4) = 2|I_1 \cup \cdots \cup I_4|$. Combining everything, we get that

$$\begin{aligned}
\mathop{\mathbb{E}}_{x}\left[g(\pi)^4\right] &\geqslant (1 - o(1)) \sum_{\substack{I_1,\ldots,I_4 \\ J_1,\ldots,J_4 \\ \text{from } \mathcal{H}}} (p(1-p))^{|I_1 \cup \cdots \cup I_4|} |a(I_1, J_1)| \cdots |a(I_4, J_4)| \\
&\geqslant (1 - o(1)) \sum_{\substack{I_1,\ldots,I_4 \\ J_1,\ldots,J_4 \\ \text{from } \mathcal{H}}} \frac{1}{n^{|I_1 \cup \cdots \cup I_4|}} |a(I_1, J_1)| \cdots |a(I_4, J_4)|. \qquad (13)
\end{aligned}$$

Combining (12) and (13) shows that $\|f\|_4^4 \leqslant (1 + o(1)) \|g\|_4^4$. $\square$

6.5 Deducing hypercontractivity for low-degree functions

With Theorem 6.8 in hand, one may deduce the following inequality as an easy corollary.

Corollary 6.12. *There exists an absolute constant $C > 0$ such that for all $d, n \in \mathbb{N}$ for which $n \geqslant 2^{C \cdot d \log d}$, the following holds. If $f \in V_d(S_n)$ is ε-analytically-global, then $\|f\|_4^4 \leqslant 2^{C \cdot d \log d} \varepsilon^2$.*

Proof. Since the proof is straightforward, we only outline its steps. Writing $f = f_0 + \cdots + f_d$ for $f_k \in V_k$ given by normalizing relations, one bounds $\|f\|_4^4 \leqslant (d+1)^3 \sum_{k=0}^{d} \|f_k\|_4^4$, uses Theorem 6.8 on each f_k, and finally $I_{I',J'}[f_k] \leqslant I_{I',J'}[f] \leqslant \varepsilon$. □

Remark 6.13. *Using the same techniques, one may prove statements analogous to Theorem 6.8 and Corollary 6.12 for all even $q \in \mathbb{N}$.*

7 Applications

7.1 Global functions are concentrated on the high degrees

The first application of our hypercontractive is the following level-d inequality.

Theorem 1.6 (Restated) . *There exists an absolute constant $C > 0$ such that the following holds. Let $d, n \in \mathbb{N}$ and $\varepsilon > 0$ such that $n \geqslant 2^{Cd^3} \log(1/\varepsilon)^{Cd}$. If $f \colon S_n \to \{0, 1\}$ is $(2d, \varepsilon)$-global, then $\|f^{\leqslant d}\|^2 \leqslant 2^{C \cdot d^4} \varepsilon^4 \log^{C \cdot d}(1/\varepsilon)$.*

Proof. Deferred to Section 8. □

This result is analogous to the level d inequality on the Boolean hypercube [24, Corollary 9.25], however it is quantitatively weaker because our dependence on d is poorer; for instance, it remains meaningful only for $d \leqslant \log(1/\varepsilon)^{1/4}$, wherein the original statement on the Boolean hypercube remains effective up to $d \sim \log(1/\varepsilon)$. Still, we show in Section 7.2 that this statement suffices to recover results regarding the size of the largest product-free sets in S_n.

It would be interesting to prove a quantitatively better version of Theorem 1.6 in terms of d, and in particular whether it is the case that for $d = c \log(1/\varepsilon)$ it holds that $\|f^{=d}\|^2 = \varepsilon^{2-o(1)}$ for sufficiently small (but constant) $c > 0$.

We remark that once Theorem 1.6 has been established (or more precisely, the slightly stronger statement in Proposition 8.11), one can strengthen it at the expense of assuming that n is larger, namely establish Theorem 1.7 from the introduction. We defer its proof to Section 8.8.

7.2 Global product-free sets are small

In this section we prove a strengthening of Theorem 1.8. Conceptually, the proof is very simple. Starting with Gowers' approach, we convert this problem into an independent set in a Cayley graph associated with F, and use a Hoffman-type bound to solve that problem.

Fix a global product-free set $F \subseteq A_n$, and construct the (directed) graph G_F as follows. Its vertex set is S_n, and (π, σ) is an edge if $\pi^{-1}\sigma \in F$. Note that G_F is a Cayley graph, and that if F is product-free, then F is an independent set in G_F. Our plan is thus to (1) study the eigenvalues of G_F and prove good upper bounds on them, and then (2) bound the size of F using a Hoffman-type bound.

Let T_F be the adjacency operator of G_F, i.e. the random walk that from a vertex π transitions to a random neighbour σ in G_F. We may consider the action of T_F on functions $f\colon S_n \to \mathbb{R}$ as

$$(T_F f)(\pi) = \mathop{\mathbb{E}}_{\sigma:(\pi,\sigma) \text{ is an edge}} [f(\sigma)] = \mathop{\mathbb{E}}_{a\in F} [f(\pi a)].$$

We will next study the eigenspaces and eigenvalues of T_F, and for that we need some basic facts regarding the representation theory of S_n. We will then study the fraction of edges between any two global functions $\mathcal{A}, \mathcal{B}$, and Theorem 1.8 will just be the special case that $\mathcal{A} = \mathcal{B} = F$.

Throughout this section, we set $\delta = \frac{|F|}{|S_n|}$.

7.2.1 Basic facts about representation theory of S_n

We will need some basic facts about the representation theory of S_n, and our exposition will follow standard textbooks, e.g. [13].

A partition of $[n]$, denoted by $\lambda \vdash n$, is a sequence of integers $\lambda = (\lambda_1, \ldots, \lambda_k)$ where $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k \geqslant 1$ sum up to n. It is well-known that partitions index equivalence classes of representations of S_n, thus we may associate with each partition λ a character $\chi_\lambda\colon S_n \to \mathbb{C}$, which in the case of the symmetric group is real-valued. The dimension of λ is $\mathsf{dim}(\lambda) = \chi_\lambda(e)$, where e is the identity permutation.

Given a partition λ, a λ-tabloid is a partition of $[n]$ into sets $A_1, \ldots, A_k$ such that $|A_i| = \lambda_i$. Thus, for λ-tabloids $A = (A_1, \ldots, A_k)$ and $B = (B_1, \ldots, B_k)$, we define $T_{A,B} = \{\pi \in S_n \mid \pi(A_i) = B_i\ \forall i = 1, \ldots, k\}$, and refer to any such $T_{A,B}$ as a λ-coset.

With these notations, we may define the space $V_\lambda(S_n)$, which is the linear span of the indicator functions of all λ-cosets. We note that $V_\lambda(S_n)$ is clearly a left S_n-module, where the action of S_n is given as ${}^\pi f\colon S_n \to \mathbb{R}$ defined by ${}^\pi f(\sigma) = f(\pi\sigma)$.

Next, we need to define an ordering on partitions that will let us further refine the spaces V_λ.

Definition 7.1. *Let $\lambda = (\lambda_1, \ldots, \lambda_k)$, $\mu = (\mu_1, \ldots, \mu_s)$ be partitions of $[n]$. We say that λ dominates μ, and denote $\lambda \trianglerighteq \mu$, if for all $j = 1, \ldots, k$ it holds that $\sum\limits_{i=1}^{j} \lambda_i \geqslant \sum\limits_{i=1}^{j} \mu_i$.*

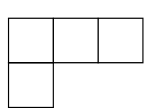
With this definition, one may easily show that $V_\mu \subseteq V_\lambda$ whenever $\mu \trianglerighteq \lambda$, and furthermore that $V_\mu = V_\lambda$ if and only if $\mu = \lambda$. It thus makes sense to define the spaces

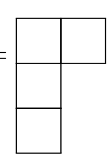
$$V_{=\lambda} = V_\lambda \cap \bigcap_{\mu \triangleright \lambda} V_\mu^\perp.$$

The spaces $V_{=\lambda}$ are orthogonal and their direct sum is $\{f\colon S_n \to \mathbb{R}\}$, so we may write any function $f\colon S_n \to \mathbb{R}$ as $f = \sum\limits_{\lambda \vdash n} f^{=\lambda}$ in a unique way.

Definition 7.2. *Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ be a partition of n. The transpose partition, λ^t, is $(\mu_1, \ldots, \mu_{k'})$, where $k' = \lambda_1$ and $\mu_j = |\{ i \mid \lambda_i \geqslant j\}|$.*

Alternatively, if we think of a partition as represented by top-left justified rows, then the transpose of a partition is obtained by reflecting the diagram across the main diagonal. For example, $(3,1)^t = (2,1,1)$:

$(3,1) =$ 

$(2,1,1) =$ 

There are two partitions that are very easy to understand: $\lambda = (n)$, and its transpose, $\lambda = (1^t)$. For $\lambda = (n)$, the space $V_{=\lambda}$ consists of constant functions, and one has $\chi_\lambda = 1$. Thus, $f^{=(n)}$ is just the average of f, i.e. $\mu(f) \overset{def}{=} \mathbb{E}_\pi[f(\pi)]$. For $\lambda = (1^n)$, the space $V_{=\lambda}$ consists of multiples of the sign function of permutations, $\mathsf{sign}\colon S_n \to \{-1,1\}$, and $\chi_\lambda = \mathsf{sign}$. One therefore has $f^{=\lambda} = \langle f, \mathsf{sign}\rangle \mathsf{sign}(f)$.

For general partitions λ, it is well-known that the dimensions of λ and λ^t are equal, and one has that $\chi_{\lambda^t} = \mathsf{sign}\cdot\chi_\lambda$. We will need the following statement that generalizes this correspondence to $f^{=\lambda}$ and $f^{=\lambda^t}$.

Lemma 7.3. *Let $f\colon S_n \to \mathbb{R}$, and let $\lambda \vdash n$. Then $(f\cdot\mathsf{sign})^{=\lambda} = f^{=\lambda^t}\mathsf{sign}$.*

Proof. The statement follows directly from the inversion formula for $f^{=\lambda}$, which states that $f^{=\lambda}(\pi) = \mathsf{dim}(\lambda)\mathbb{E}_{\sigma\in S_n}\left[f(\sigma)\chi_\lambda(\pi\sigma^{-1})\right]$. By change of variables, we see that

$$(f\cdot\mathsf{sign})^{=\lambda}(\pi) = \mathsf{dim}(\lambda)\mathop{\mathbb{E}}_{\sigma\in S_n}\left[f(\sigma^{-1}\pi)\mathsf{sign}(\sigma^{-1}\pi)\chi_\lambda(\sigma)\right] = \mathsf{sign}(\pi)\mathsf{dim}(\lambda)\mathop{\mathbb{E}}_{\sigma\in S_n}\left[f(\sigma^{-1}\pi)\mathsf{sign}(\sigma)\chi_\lambda(\sigma)\right],$$

where we used the fact that sign is multiplicative and $\mathsf{sign}(\sigma^{-1}) = \mathsf{sign}(\sigma)$. Now, as $\mathsf{sign}(\sigma)\chi_\lambda(\sigma) = \chi_{\lambda^t}(\sigma)$, we get by changing variables again that

$$(f\cdot\mathsf{sign})^{=\lambda}(\pi) = \mathsf{sign}(\pi)\mathsf{dim}(\lambda)\mathop{\mathbb{E}}_{\sigma\in S_n}\left[f(\sigma)\chi_{\lambda^t}(\pi\sigma^{-1})\right] = \mathsf{sign}(\pi)\mathsf{dim}(\lambda^t)\mathop{\mathbb{E}}_{\sigma\in S_n}\left[f(\sigma)\chi_{\lambda^t}(\pi\sigma^{-1})\right],$$

which is equal to $\mathsf{sign}(\pi)f^{=\lambda^t}(\pi)$ by the inversion formula. $\square$

Lastly, we remark that if λ is a partition such that $\lambda = n-k$, then $V_{=\lambda} \subseteq V_k$. It follows by Parseval that

$$\sum_{\substack{\lambda\vdash n\\ \lambda_1=n-k}} \left\|f^{=\lambda}\right\|_2^2 \leqslant \left\|f^{\leqslant k}\right\|_2^2. \tag{14}$$

7.2.2 The eigenvalues of T_F

Claim 7.4. *For all $\lambda \vdash n$ we have that $T_F V_{=\lambda} \subseteq V_{=\lambda}$; the same holds for T_F^*.*

Proof. First, we show that $T_F V_\lambda \subseteq V_\lambda$, and for that it is enough to show that $T_F 1_{T_{A,B}} \in V_\lambda$ for all λ-tabloids $A = (A_1,\ldots,A_k)$ and $B = (B_1,\ldots,B_k)$. Fix $a \in F$, and note that $1_{T_{A,B}}(\sigma a) = 1_{T_{a(A),B}}(\sigma)$ where $a(A) = (a(A_1),\ldots,a(A_k))$, so $1_{T_{A,B}}(\sigma a)$, as a function of σ, is also an indicator of a λ-coset. Since $T_F 1_{T_{A,B}}$ is a linear combination of such functions, it follows that $T_F 1_{T_{A,B}} \in V_\lambda$. A similar argument shows that the same holds for the adjoint operator of T_F^* (which is nothing but $T_{F^{-1}}$, where $F^{-1} = \left\{a^{-1} \;\middle|\; a\in F\right\}$).

Thus, for $f \in V_{=\lambda}$ we automatically have that $f \in V_\lambda$, and we next show orthogonality to V_μ for all $\mu \rhd \lambda$. Indeed, let μ be such partition and let $g \in V_\mu$; then by the above $T_F^* g \in V_\mu$ and so $\langle T_F f, g\rangle = \langle f, T_F^* g\rangle = 0$, and the proof is complete. The argument for T_F^* is analogous. $\square$

Thus, we may find a basis of each $V_{=\lambda}$ consisting of eigenvectors of $T_F^* T_F$. The following claim shows that the multiplicity of each corresponding eigenvalue is at least $\mathsf{dim}(\lambda)$.

Claim 7.5. *Let $f \in V_{=\lambda}(S_n)$ be non-zero. Then $\mathsf{dim}(\mathsf{Span}(\{{}^\pi f\}_{\pi\in S_n})) \geqslant \mathsf{dim}(\lambda)$.*

Proof. Let $\rho_\lambda\colon S_n \to V_{=\lambda}$ be a representation, and denote by W the span of $\{{}^\pi f\}_{\pi\in S_n}$. Note that W is a subspace of $V_{=\lambda}$, and it holds that $(\rho|_W, W)$ is a sub-representation of ρ. Since each irreducible representation $V \subseteq V_{=\lambda}$ of S_n has dimension $\mathsf{dim}(\lambda)$, it follows that $\mathsf{dim}(W) \geqslant \mathsf{dim}(\lambda)$, and we're done. $\square$

We can thus use the trace method to bound the magnitude of each eigenvalue.

Lemma 7.6. *Let $f \in V_{=\lambda}$ be an eigenvector of $T_F^* T_F$ with eigenvalue α_λ. Then*

$$\alpha_\lambda \leqslant \frac{1}{\mathsf{dim}(\lambda)\delta}.$$

Proof. By Claim 7.5, we may find a collection of $\mathsf{dim}(\lambda)$ permutations, call it Π, such that $\{{}^{\pi} f\}_{\pi\in\Pi}$ is linearly independent. Since f is an eigenvector of $T_F^* T_F$, it follows that each one of ${}^{\pi} f$ is an eigenvector with eigenvalue α_λ. It follows that $\mathsf{Tr}(T_F^* T_F) \geqslant |\Pi|\,\alpha_\lambda = \mathsf{dim}(\lambda)\alpha_\lambda$.

On the other hand, interpreting $\mathsf{Tr}(T_F^* T_F)$ probabilistically as the probability to return to the starting vertex in 2-steps,

$$\mathsf{Tr}(T_F^* T_F) = \sum_{\pi} \Pr_{a_1\in F^{-1}, a_2\in F}[\pi = \pi a_1 a_2] = n! \Pr_{a_1\in F^{-1}, a_2\in F}\left[a_2 = a_1^{-1}\right] = n!\frac{1}{|F|} = \frac{1}{\delta}.$$

Combining the two bounds on $\mathsf{Tr}(T_F^* T_F)$ completes the proof. □

To use this lemma effectively, we have the following bound on $\mathsf{dim}(\lambda)$ that follows from the hook length formula.

Lemma 7.7 (Claim 1, Theorem 19 in [7])**.** *Let $\lambda \vdash n$ be given as $\lambda = (\lambda_1, \ldots, \lambda_k)$, and denote $d = \min(n - \lambda_1, k)$.*

1. *If $\lambda = (n)$, then* $\mathsf{dim}(\lambda) = 1$.
2. *If $d > 0$, then* $\mathsf{dim}(\lambda) \geqslant \left(\frac{n}{d\cdot e}\right)^d$.
3. *If $d > n/10$, then* $\mathsf{dim}(\lambda) \geqslant 1.05^n$.

7.2.3 Applying Hoffman's bound

With the information we have gathered regarding the representation theory of S_n and the eigenvalues of T_F, we can use the spectral method to prove lower bounds on $\langle T_F g, h\rangle$ for Boolean functions g, h that are global, as in the following lemma.

Lemma 7.8. *There exists $C > 0$ such that the following holds. Let $n \in \mathbb{N}$ and $\varepsilon > 0$ be such that $n \geqslant \log(1/\varepsilon)^C$, and suppose that $g, h \colon A_n \to \{0,1\}$ are $(6,\varepsilon)$-global. Then*

$$\langle T_F g, h\rangle \geqslant \frac{\mathbb{E}[g]\,\mathbb{E}[h]}{4} - C\frac{\varepsilon^4 \log^C(1/\varepsilon)}{\sqrt{n\delta}} - \frac{1}{\sqrt{n^4\delta}}\sqrt{\frac{\mathbb{E}[g]\,\mathbb{E}[h]}{4}}.$$

Proof. Extend g, h to S_n by defining them to be 0 outside A_n.

Recall that T_F preserves each $V_{=\lambda}$. Decomposing $g = \sum_{\lambda\vdash n} g^{=\lambda}$ where $g^{=\lambda} \in V_{=\lambda}$ and h similarly, we have by Plancherel that $\langle T_F g, h\rangle = \sum\limits_{\lambda,\theta} \langle T_F g^{=\lambda}, h^{=\lambda}\rangle$. For the trivial partition $\lambda = (n)$ we have that $g^{=\lambda} \equiv \mu(g) = \mathbb{E}[g]/2$, $h^{=\lambda} \equiv \mu(h) = \mathbb{E}[h]/2$. For $\lambda = (1^n)$, since $F \subseteq A_n$ it follows that $T_F\mathsf{sign} = \mathsf{sign}$, and so $g^{=\lambda} = \beta_\lambda\mathsf{sign}$, $h^{=\lambda} = \gamma_\lambda\mathsf{sign}$ for $\beta_\lambda, \gamma_\lambda \geqslant 0$, so the term corresponding to λ in the above is non-negative. Thus, denoting $\lambda = (\lambda_1, \ldots, \lambda_k)$ we have that

$$\langle T_F g, h\rangle \geqslant \mu(g)\mu(h) - \sum_{\substack{\lambda\vdash n \\ \lambda\neq(n),(1^n) \\ \lambda_1\geqslant n-3 \text{ or } k\geqslant n-3}} \left\|T_F g^{=\lambda}\right\|_2 \left\|h^{=\lambda}\right\|_2 - \sum_{\substack{\lambda\neq(n),(1^n) \\ \lambda_1\leqslant n-4 \text{ and } k\leqslant n-4}} \left\|T_F g^{=\lambda}\right\|_2 \left\|h^{=\lambda}\right\|_2. \quad (15)$$

We upper-bound the second and third terms on the right-hand side, from which the lemma follows. We begin with the second term, and handle separately λ's such that $\lambda_1 \geqslant n-3$, and λ's such that $k \geqslant n-3$.

λ's such that $\lambda_1 \geqslant n-3$. We first upper bound $\left\|T_F g^{=\lambda}\right\|_2$. As $T_F^* T_F$ preserves each space $V_{=\lambda}$ and is symmetric, we may write this space as a sum of eigenspaces of $T_F^* T_F$, say $\bigoplus_\theta V_{=\lambda}^\theta$. We thus write Writing $g^{=\lambda} = \sum_\theta g^{=\lambda,\theta}$ where $g^{=\lambda,\theta} \in V_{=\lambda}^\theta$, and have that

$$\left\|T_F g^{=\lambda}\right\|_2^2 = \langle g^{=\lambda}, T_F^* T_F g^{=\lambda}\rangle = \sum_\theta \langle g^{=\lambda,\theta}, T_F^* T_F g^{=\lambda,\theta}\rangle = \sum_\theta \theta \left\|g^{=\lambda,\theta}\right\|_2^2$$

By Lemma 7.6 we have $\theta \leqslant \frac{1}{\mathsf{dim}(\lambda)\delta}$, which by Fact 7.7 is at most $O\left(\frac{1}{n\delta}\right)$. We thus get that

$$\left\|T_F g^{=\lambda}\right\|_2^2 \leqslant O\left(\frac{1}{n\delta}\right) \sum_\theta \left\|g^{=\lambda,\theta}\right\|_2^2 \leqslant O\left(\frac{1}{n\delta}\right) \left\|g^{=\lambda}\right\|_2^2$$

Plugging this into the second sum in (15), we get that the second sum there is at most

$$O\left(\frac{1}{n\delta}\right) \sum_{\substack{\lambda\vdash n \\ \lambda\neq(n),(1^n) \\ \lambda_1\geqslant n-3}} \left\|g^{=\lambda}\right\|_2 \left\|h^{=\lambda}\right\|_2 \leqslant O\left(\frac{1}{n\delta}\right) \left\|g^{\leqslant 3}\right\|_2 \left\|h^{\leqslant 3}\right\|_2,$$

where we used Cauchy-Schwarz and (14). By Theorem 1.6, $\left\|g^{\leqslant 3}\right\|_2^2, \left\|h^{\leqslant 3}\right\|_2^2 \leqslant C \cdot \varepsilon^4 \log^C(1/\varepsilon)$ for some absolute constant C. We thus get that

$$\sum_{\substack{\lambda\vdash n \\ \lambda\neq(n),(1^n) \\ \lambda_1\geqslant n-3}} \left\|T_F g^{=\lambda}\right\|_2 \left\|h^{=\lambda}\right\|_2 \leqslant \frac{1}{\sqrt{n\delta}} C' \cdot \varepsilon^4 \log^C(1/\varepsilon).$$

λ's such that $k \geqslant n-3$. The treatment here is pretty much identical to the previous case, except that we look at the functions $\tilde{g} = g \cdot \mathsf{sign}$ and $\tilde{h} = h \cdot \mathsf{sign}$. That is, first note that the globalness of g, h implies that $\tilde{g}, \tilde{h}$ are also global with the same parameters, and since g, h are Boolean, $\tilde{g}, \tilde{h}$ are integer valued. Moreover, by Lemma 7.3 we have that

$$\sum_{\substack{\lambda\vdash n \\ \lambda\neq(n),(1^n) \\ k\geqslant n-3}} \left\|T_F g^{=\lambda}\right\|_2 \left\|h^{=\lambda}\right\|_2 = \sum_{\substack{\lambda\vdash n \\ \lambda\neq(n),(1^n) \\ k\geqslant n-3}} \left\|T_F \tilde{g}^{=\lambda^t}\right\|_2 \left\|\tilde{h}^{=\lambda^t}\right\|_2 = \sum_{\substack{\lambda\vdash n \\ \lambda\neq(n),(1^n) \\ \lambda_1\geqslant n-3}} \left\|T_F \tilde{g}^{\lambda}\right\|_2 \left\|\tilde{h}^{\lambda}\right\|_2,$$

and from here the argument is identical.

Bounding the third term in (15). Repeating the eigenspace argument from above, we have that for all $\lambda \vdash n$ for which $\lambda_1 \leqslant n-4$ and $k \leqslant n-4$ we have that

$$\left\|T_F g^{=\lambda}\right\|_2 \leqslant O\left(\frac{1}{\sqrt{n^4\delta}}\right) \left\|g^{=\lambda}\right\|_2.$$

Thus, the third sum is bounded by

$$O\left(\frac{1}{\sqrt{n^4\delta}}\right) \sum_{\lambda\vdash n} \left\|g^{=\lambda}\right\|_2 \left\|h^{=\lambda}\right\|_2 \leqslant O\left(\frac{1}{\sqrt{n^4\delta}}\right) \|g\|_2 \|h\|_2.$$

where we used Cauchy–Schwarz and Parseval. $\square$

We can now prove the strengthening of Theorem 1.8, stated below.

Corollary 7.9. *There exists $K \in \mathbb{N}$ such that the following holds for all $\varepsilon > 0$ and $n \geqslant \log^K(1/\varepsilon)$. If $\mathcal{A}, \mathcal{B} \subseteq A_n$ are $(6, \varepsilon)$-global, and $\mu(\mathcal{A})\mu(\mathcal{B}) \geqslant K \max(n^{-4}\delta^{-1}, (n\delta)^{-1/2}\varepsilon^4 \log^K(1/\varepsilon))$, then*

$$\langle T_F g, h\rangle \geqslant \frac{1}{5}\mu(\mathcal{A})\mu(\mathcal{B}).$$

Proof. Taking $g = 1_{\mathcal{A}}$, $h = 1_{\mathcal{B}}$, by Lemma 7.8 we have

$$\langle T_F g, h\rangle \geqslant \frac{1}{4}\mu(\mathcal{A})\mu(\mathcal{B}) - C' \frac{\varepsilon^4 \log^{C'}(1/\varepsilon)}{\sqrt{n\delta}} - \frac{2}{n^2\sqrt{\delta}}\sqrt{\mu(\mathcal{A})\mu(\mathcal{B})},$$

where C' is an absolute constants. Now the conditions on the parameters implies that the first term dominates the other two. □

We note that Theorem 1.8 immediately follows, since there one has $g = h = 1_F$ and $\langle T_F g, h\rangle = 0$, so one gets that the condition on the parameters fail, and therefore the lower bound on $\mu(\mathcal{A})\mu(\mathcal{B})$ (which in this case is just δ^2) fails; plugging in $\varepsilon = C \cdot \sqrt{\delta}$ and rearranging finishes the proof.

7.2.4 Improving on Theorem 1.8?

We remark that it is within reason to expect that global, product-free families in A_n must in fact be much smaller. To be more precise, one may expect that for all $t \in \mathbb{N}$, there is $j \in \mathbb{N}$ such that for $n \geqslant n_0(t)$, if F is $(j, O(\sqrt{\delta}))$-global (where $\delta = |F| / |S_n|$), then $\delta \leqslant O_t(n^{-t})$. The bottleneck in our approach comes from the use of the trace method (which doesn't use the globalness of F at all), and the bounds it gives on the eigenvalues of $T_F^* T_F$ corresponding to low-degree functions: they become meaningless as soon as $\delta \geqslant 1/n$.

Inspecting the above proof, our approach only requires a super-logarithmic upper bound on the eigenvalues to go through. More precisely, we need that the first few non-trivial eigenvalues of $T_F^* T_F$ are at most $(\log n)^{-K(t)}$, for sufficiently large $K(t)$. We feel that something like that should follow in greater generality from the fact that the set of generators in the Cayley graph, namely F, is global. To support that, note that if we were dealing with Abelian groups, then the eigenvalue α of T_F corresponding to a character χ could be computed as $\lambda = \frac{1}{|F|}\sum\limits_{a \in F} \chi(a)$, which by rewriting is nothing but a (normalized) Fourier coefficient of F, i.e. $\frac{1}{\delta}\widehat{1_F}(\chi)$, which we expect to be small by the globalness of F.

7.3 Isoperimetric inequalities in the transpositions Cayley graph

In this section, we consider T which is the adjacency operator of the transpositions graph. That is, it is the transition matrix of the (left) Cayley graph (S_n, A), where A is the set of transpositions (and the multiplication happens from the left). We show that for a global set S, starting a walk from a vertex in S and performing $\approx cn$ steps according to T escapes S with probability close to 1.

Poisson process random walk. To be more precise, we consider the following random walk: from a permutation $\pi \in S$, choose a number $k \sim \mathsf{Poisson}(t)$, take τ which is a product of k random transpositions, and go to $\sigma = \tau \circ \pi$. We show that starting with a random $\pi \in S$, the probability that we escape S, i.e. that $S\sigma \notin S$, is close to 1.

To prove this result, we first note that the distribution of an outgoing neighbour from π is exactly $e^{-t(I-\mathrm{T})}1_\pi$, where 1_π is the indicator vector of π. Therefore, the distribution of σ where $\pi \in S$ is random is $e^{-t(I-\mathrm{T})}\frac{1_S}{|S|}$, where 1_S is the indicator vector of S. Thus, the probability that σ is in S (i.e. of the complement event) is

$$\frac{1}{\mu(S)}\langle 1_S, e^{-t(I-\mathrm{T})}1_S\rangle,$$

where $\mu(S)$ is the measure of S. We upper-bound this quantity using spectral considerations. We will only need our hypercontractive inequality and basic knowledge of the eigenvalues of T, which can be found, for example, in [10, Corollary 21]. This is the content of the firs three items in the lemma below (we also prove a fourth item, which will be useful for us later on).

Lemma 7.10. *Let $\lambda \in \mathbb{R}$ be an eigenvalue of T, and $f \in V_d(S_n)$ be a corresponding eigenvector.*

1. $\mathrm{T}V_{=d}(S_n) \subseteq V_{=d}(S_n)$.
2. $1 - \frac{2d}{n-1} \leqslant \lambda \leqslant 1 - \frac{d}{n-1}$.
3. *If $d \leqslant n/2$, then we have the stronger bound* $1 - \frac{2d}{n-1} \leqslant \lambda \leqslant 1 - \left(1 - \frac{d-1}{n}\right)\frac{2d}{n-1}$.
4. *If L is a Laplacian of order 1, then L and T commute. Thus, T commutes with all Laplacians.*

Proof. For the first item, we first note that T commutes with the right action of S_n on functions:

$$(\mathrm{T}(f^\pi))(\sigma) = \mathop{\mathbb{E}}_{\pi' \text{ a transposition}}\left[f^\pi(\pi' \circ \sigma)\right] = \mathop{\mathbb{E}}_{\pi' \text{ a transposition}}\left[f(\pi' \circ \sigma \circ \pi)\right] = \mathrm{T}f(\sigma \circ \pi) = (\mathrm{T}f)^\pi(\sigma).$$

Also, T is self adjoint, so T^* also commutes with the action of S_n. The first item now follows as in the proof of Claim 7.4.

The second and third items are exactly [10, Corollary 21]. For the last item, for any function f and an order 1 Laplacian $\mathrm{L} = \mathrm{L}_{(i,j)}$,

$$\mathrm{T}\mathrm{L}f = \mathrm{T}\left(f - f^{(i,j)}\right) = \mathrm{T}f - \mathrm{T}\left(f^{(i,j)}\right) = \mathrm{T}f - (\mathrm{T}f)^{(i,j)} = \mathrm{L}\left(\mathrm{T}f\right),$$

where in the third transition we used the fact that T commutes with the right action of S_n. $\square$

We remark that the first item above implies that we may find a basis of the space of real-valued functions consisting of eigenvectors of T, where each function is from $V_{=d}(S_n)$ for some d. Lastly, we need the following (straightforward) fact.

Fact 7.11. *If $f \in V_d(S_n)$ is an eigenvector of T with eigenvalue λ, then f is an eigenvector of $e^{-t(I-\mathrm{T})}$ with eigenvalue $e^{-t(1-\lambda)}$.*

Theorem 7.12. *There exists $C > 0$ such that the following holds for all $d \in \mathbb{N}$, $t, \varepsilon > 0$ and $n \in \mathbb{N}$ such that $n \geqslant 2^{C \cdot d^3}\log^{C\cdot d}(1/\varepsilon)$. If $S \subseteq S_n$ is a set of vertices such that 1_S is $(2d, \varepsilon)$-global, then*

$$\Pr_{\substack{\pi \in S \\ \sigma \sim e^{-t(I-T)}\pi}}[\sigma \notin S] \geqslant 1 - \left(2^{C\cdot d^4}\varepsilon \log^{C\cdot d}(1/\varepsilon) + e^{-\frac{(d+1)t}{n-1}}\right).$$

Proof. Consider the complement event that $\sigma \in S$, and note that the desired probability can be written analytically as $\frac{1}{\mu(S)}\langle 1_S, e^{-t(I-\mathrm{T})}1_S\rangle$, where $\mu(S)$ is the measure of S. Now, writing $f = 1_S$ and expanding $f = f_{=0} + f_{=1} + \cdots$, we consider each one of $e^{-t(I-\mathrm{T})}f^{=j}$ separately. We claim that

$$\left\|e^{-t(I-\mathrm{T})}f^{=j}\right\|_2 \leqslant e^{-\frac{jt}{n-1}}\left\|f^{=j}\right\|_2. \tag{16}$$

Indeed, note that we may write $f^{=j} = \sum a_r f_{j,r}$, where $f_{j,r} \in V_{=j}(S_n)$ are orthogonal and eigenvectors of T with eigenvalue $\lambda_{j,r}$, and so by Fact 7.11, $e^{-t(I-\mathrm{T})}f^{=j} = \sum\limits_r e^{-t(1-\lambda_{j,r})}f_{j,r}$. By Parseval we deduce that

$$\left\|e^{-t(I-\mathrm{T})}f^{=j}\right\|_2^2 \leqslant \sum_r e^{-t(1-\lambda_{j,r})}\left\|f_{j,r}\right\|_2^2 \leqslant \max_r e^{-t(1-\lambda_{j,r})}\sum_r \left\|f_{j,r}\right\|_2^2 = \max_r e^{-t(1-\lambda_{j,r})}\left\|f^{=j}\right\|_2^2.$$

Inequality (16) now follows from the second item in Lemma 7.10.

We now expand out the expression we have for the probability of the complement event using Plancherel:

$$\begin{aligned}\frac{1}{\mu(S)}\langle 1_S, e^{-t(I-\mathrm{T})}1_S\rangle = \frac{1}{\mu(S)}\sum_j \langle f^{=j}, e^{-t(I-\mathrm{T})}f^{=j}\rangle &\leqslant \frac{1}{\mu(S)}\sum_j \left\|f^{=j}\right\|_2\left\|e^{-t(I-\mathrm{T})}f^{=j}\right\|_2 \\ &\leqslant \frac{1}{\mu(S)}\sum_j e^{-\frac{jt}{n-1}}\left\|f^{=j}\right\|_2^2, \end{aligned} \tag{17}$$

where in the last two transitions we used Cauchy–Schwarz and inequality (16). Lastly, we bound $\left\|f^{=j}\right\|_2^2$. For $j > d$ we have that $\sum\limits_{j>d}\left\|f^{=j}\right\|_2^2 \leqslant \mu(S)$ by Parseval, and for $j \leqslant d$ we use hypercontractivity.

First, bound $\left\|f^{=j}\right\|_2 \leqslant \left\|f^{\leqslant j}\right\|_2$, and note that the function $f^{\leqslant j}$ is $(2j, 2^{O(j^4)}\varepsilon^2\log^{O(j)}(1/\varepsilon))$-global by Claim A.1. Thus, using Hölder's inequality and Theorem 1.4 we get that

$$\left\|f^{\leqslant j}\right\|_2^2 = \langle f, f^{\leqslant j}\rangle \leqslant \|f\|_{4/3}\left\|f^{\leqslant j}\right\|_4 \leqslant \mu(S)^{3/4}2^{O(j^3)}\sqrt{2^{O(j^4)}\varepsilon^2\log^{O(j)}(1/\varepsilon)}\left\|f^{\leqslant j}\right\|_2^{1/2}.$$

Rearranging gives $\left\|f^{\leqslant j}\right\|_2^2 \leqslant 2^{O(j^4)}\mu(S)\varepsilon\log^{O(j)}(1/\varepsilon)$.

Plugging our estimates into (17) we get

$$\frac{1}{\mu(S)}\langle 1_S, e^{-t(I-\mathrm{T})}1_S\rangle \leqslant \sum_{j=0}^{d} 2^{O(j^4)}e^{-\frac{jt}{n-1}}\varepsilon\log^{O(j)}(1/\varepsilon) + e^{-\frac{(d+1)t}{n-1}} \leqslant 2^{O(d^4)}\varepsilon\log^{O(d)}(1/\varepsilon) + e^{-\frac{(d+1)t}{n-1}}.$$

□

Using exactly the same technique, one can prove a lower bound on the probability of escaping a global set in a single step, as stated below. This result is similar in spirit to a variant of the KKL Theorem over the Boolean hypercube [15], and therefore we modify the formulation slightly. Given a function $f\colon S_n \to \mathbb{R}$, we define the influence of coordinate $i \in [n]$ to be

$$I_i[f] = \mathop{\mathbb{E}}_{j\neq i}\left[\left\|L_{(i,j)}f\right\|_2^2\right],$$

and define the total influence of f to be $I[f] = I_1[f] + \cdots + I_n[f]$.

Theorem 7.13. *There exists $C > 0$ such that the following holds for all $d \in \mathbb{N}$ and $n \in \mathbb{N}$ such that $n \geqslant 2^{C \cdot d^3}$. Suppose $S \subseteq S_n$ is such that for all derivative operators $\mathrm{D} \neq I$ of order at most d, it holds that $\|\mathrm{D}1_S\|_2 \leqslant 2^{-C \cdot d^4}$. Then*

$$I[1_S] \geqslant \frac{1}{4} d \cdot \mathsf{var}(1_S).$$

Proof. Deferred to Appendix A. □

7.4 Deducing results for the multi-cube

Our hypercontractive inequalities also imply similar hypercontractive inequalities on different non-product domains. One example from [3] is the domain of 2-to-1 maps, i.e. $\left\{\pi : [2n] \to [n] \mid \left|\pi^{-1}(i)\right| = 2\ \forall i \in [n]\right\}$. A more general domain, which we consider below, is the multi-slice.

Definition 7.14. *Let $m, n \in \mathbb{N}$ such that $n \geqslant m$, and let $k_1, \ldots, k_m \in \mathbb{N}$ sum up to n. The multi-slice $\mathcal{U}_{k_1,\ldots,k_m}$ of dimension n consists of all vectors $x \in [m]^n$ that, for all $j \in [m]$, have exactly k_j of their coordinates equal to j.*

We consider the multi-slice as a probability space with the uniform measure.

In exactly the same way one defines the degree decomposition over S_n, one may consider the degree decomposition over the mutli-slice. A function $f : \mathcal{U}_{k_1,\ldots,k_m} \to \mathbb{R}$ is said to be a d-junta if there are $A \subseteq [n]$ of size at most d and $g : [m]^d \to \mathbb{R}$ such that $f(x) = g(x_A)$. We then define the space $V_d(\mathcal{U}_{k_1,\ldots,k_m})$ spanned by d-juntas. Also, one may analogously define globalness of functions over the multi-slice. A d-restriction consists of a set $A \subseteq [n]$ of size d and $\alpha \in [m]^A$, and the corresponding restriction is the function $f_{A \to \alpha}(z) = f(x_A = \alpha, x_{\bar{A}} = z)$ (whose domain is a different multi-slice).

Definition 7.15. *We say $f : \mathcal{U}_{k_1,\ldots,k_m} \to \mathbb{R}$ is (d, ε)-global if for any d-restriction (A, α) it holds that $\|f_{A \to \alpha}\|_2 \leqslant \varepsilon$.*

7.4.1 Hypercontractivity

Our hypercontractive inequality for the multi-slice reads as follows.

Theorem 7.16. *There exists an absolute constant $C > 0$ such that the following holds. Let $d, q, n \in \mathbb{N}$ be such that $n \geqslant q^{C \cdot d^2}$, and let $f \in V_d(\mathcal{U}_{k_1,\ldots,k_m})$. If f is $(2d, \varepsilon)$-global, then*

$$\|f\|_q \leqslant q^{O\left(d^3\right)} \varepsilon^{\frac{q-2}{q}} \|f\|_2^{\frac{2}{q}}.$$

Proof. We construct a simple deterministic coupling $\mathcal{C}$ between S_n and $\mathcal{U}_{k_1,\ldots,k_m}$.

Fix a partition of $[n]$ into sets $K_1, \ldots, K_m$ such that $|K_j| = k_j$ for all j. Given a permutation π, we define $\mathcal{C}(\pi) = x$ as follows: for all $i \in [n]$, $j \in [m]$, we set $x_i = j$ if $\pi(i) \in K_j$. Define the mapping $M : L_2(\mathcal{U}_{k_1,\ldots,k_m}) \to L_2(S_n)$ that maps a function $h : \mathcal{U}_{k_1,\ldots,k_m} \to \mathbb{R}$ to $Mh : S_n \to \mathbb{R}$ defined by $(Mh)(\pi) = h(\mathcal{C}(\pi))$.

Let $g = Mf$. We claim that g has degree at most d and is global. To see that $g \in V_d(S_n)$, it is enough to show that the mapping $f \to g$ is linear (which is clear), and maps a d-junta into a d-junta, which is also straightforward. To see that g is global, let $T = \{(i_1, r_1), \ldots, (i_\ell, r_\ell)\}$ be consistent, and define the r-restriction (A, α) as: $A = \{i_1, \ldots, i_\ell\}$, and $\alpha_{i_s} = j$ if $r_s \in K_j$. Note that the distribution of $x \in \mathcal{U}_{k_1,\ldots,k_m}$ conditioned on x_A is exactly the same as of $\mathcal{C}(\pi)$ conditioned on π respecting T, so if $r \leqslant 2d$ we get that

$$\|g_{A \to \alpha}\|_2 = \|f_{\to T}\|_2 \leqslant \varepsilon,$$

and g is $(2d, \varepsilon)$-global. The result thus follows from Theorem 1.4 and the fact that M preserves L_p norms for all $p \geqslant 1$. $\square$

The coupling in the proof of Theorem 7.16 also implies in the same way a level-d inequality over $\mathcal{U}_{k_1,\dots,k_m}$ from the corresponding result in S_n, Theorem 1.6, as well as isoperimetric inequalities, as we describe next.

7.4.2 Level-d inequality

As on S_n, for $f\colon \mathcal{U}_{k_1,\dots,k_m} \to \mathbb{R}$ we let $f^{\leqslant d}$ be the projection of f onto $V_d(\mathcal{U}_{k_1,\dots,k_m})$. Our level-$d$ inequality for the multi-slice thus reads:

Corollary 7.17. *There exists an absolute constant $C > 0$ such that the following holds. Let $d, n \in \mathbb{N}$ and $\varepsilon > 0$ such that $n \geqslant 2^{Cd^3} \log(1/\varepsilon)^{Cd}$. If $f\colon \mathcal{U}_{k_1,\dots,k_m} \to \{0,1\}$ is $(2d, \varepsilon)$-global, then $\left\|f^{\leqslant d}\right\|_2^2 \leqslant 2^{C\cdot d^4} \varepsilon^4 \log^{C\cdot d}(1/\varepsilon)$.*

Proof. The proof relies on an additional easy property of the mapping M from the proof of Theorem 7.16. As in S_n, we define the space of pure degree d functions over $\mathcal{U}_{k_1,\dots,k_m}$ as $V_{=d}(\mathcal{U}_{k_1,\dots,k_m}) = V_d(\mathcal{U}_{k_1,\dots,k_m}) \cap V_{d-1}(\mathcal{U}_{k_1,\dots,k_m})^{\perp}$, and let $f^{=d}$ be the projection of f onto $V_{=d}(\mathcal{U}_{k_1,\dots,k_m})$. We thus have $f^{\leqslant d} = f^{=0} + f^{=1} + \cdots + f^{=d}$, and so $f^{=d} = f^{\leqslant d} - f^{\leqslant d-1}$.

Write $h_i = Mf^{=i}$, and note that h_i is of degree at most i. Also, we note that as restrictions of size $r < i$ over S_n are mapped to restrictions of size r over $\mathcal{U}_{k_1,\dots,k_m}$, it follows that h_i is perpendicular to degree $i-1$ functions, and so $h_i \in V_{=i}(S_n)$. By linearity of M, $Mf = h_0 + h_1 + \cdots + h_n$, and by uniqueness of the pure degree decomposition, it follows that $h_i = (Mf)^{=i}$. We therefore have that

$$\left\|f^{\leqslant d}\right\|_2^2 = \sum_{i\leqslant d} \left\|f^{=i}\right\|_2^2 = \sum_{i\leqslant d} \|h_i\|_2^2 = \sum_{i\leqslant d} \left\|(Mf)^{=i}\right\|_2^2 = \left\|(Mf)^{\leqslant d}\right\|_2^2 \leqslant 2^{C\cdot d^4} \varepsilon^4 \log^{C\cdot d}(1/\varepsilon),$$

where the last inequality is by Theorem 1.6. $\square$

7.4.3 Isoperimetric inequalities

One can also deduce the obvious analogs of Theorems 7.12, 7.13 for the multi-slice. Since we use it for our final application, we include here the statement of the analog of Theorem 7.13.

For $f\colon \mathcal{U}_{k_1,\dots,k_m} \to \mathbb{R}$, consider the Laplacians $\mathrm{L}_{i,j}$ that map a function f to a function $L_{i,j} f$ defined as $\mathrm{L}_{i,j} f(x) = f(x) - f(x^{(i,j)})$, and define $I_i[f] = \mathbb{E}_{j\neq i}\left[\|L_{i,j} f\|_2^2\right]$ and $I[f] = \sum\limits_{i=1}^{n} I_i[f]$. Similarly to Definition 4.1, we define a derivative of f as a restriction of the corresponding Laplacian, i.e. for $i, j \in [n]$, $a, b \in [m]$ we define $\mathrm{D}_{(i,j)\to(a,b)} f = (\mathrm{L}_{i,j} f(x))_{(i,j)\to(a,b)}$.

Theorem 7.18. *There exists $C > 0$ such that the following holds for all $d \in \mathbb{N}$ and $n \in \mathbb{N}$ such that $n \geqslant 2^{C\cdot d^3}$. Suppose $S \subseteq \mathcal{U}_{k_1,\dots,k_m}$ such that for all derivative operators $\mathrm{D} \neq I$ of order at most d it holds that $\|\mathrm{D}1_S\|_2 \leqslant 2^{-C\cdot d^4}$. Then $I[1_S] \geqslant \frac{1}{4} d \cdot \mathsf{var}(1_S)$.*

We omit the straightforward derivation from Theorem 7.13.

7.5 Stability result for the Kruskal–Katona theorem on the slice

Our final application is the following sharp threshold result for the slice, which can be also seen as a stability version of the Kruskal–Katona theorem (see [25, 16] for other, incomparable stability versions). For a family of subsets $\mathcal{F} \subseteq \binom{[n]}{k}$, we denote $\mu(\mathcal{F}) = |\mathcal{F}| / \binom{n}{k}$. and define the upper shadow of $\mathcal{F}$ as

$$\mathcal{F}\uparrow = \left\{ X \in \binom{n}{k+1} \,\middle|\, \exists A \subseteq X, A \in \mathcal{F} \right\}.$$

The Kruskal–Katona theorem is a basic result in combinatorics that gives a lower bound on the measure of the upper shadow of a family $\mathcal{F}$ in terms of the measure of the family itself. Below we state a convenient, simplified version of it due to Lovász', which uses the generalized binomial coefficients.

Theorem 7.19. *Let $\mathcal{F} \subseteq \binom{[n]}{k}$ and suppose that $|\mathcal{F}| = \binom{n}{x}$. Then $|\mathcal{F}\uparrow| \geqslant \binom{n}{x+1}$.*

In general, Theorem 7.19 is tight, as can be shown by considering "subcubes", i.e. families of the form $\mathcal{H} = \left\{ X \in \binom{[n]}{k} \,\middle|\, X \supseteq A \right\}$ for some $A \subseteq [n]$. This raises the question of whether a stronger version of Theorem 7.19 holds for families that are "far from having a structure such as $\mathcal{H}$". Alternatively, this question can be viewed as a stability version of Theorem 7.19: must a family for which Theorem 7.19 is almost tight be of a similar structure to $\mathcal{H}$?

Below, we mainly consider the case that $k = o(n)$, and show in improved version of Theorem 7.19 for families that are "far from $\mathcal{H}$". To formalize this, we consider the notion of restrictions: for $A \subseteq I \subseteq [n]$, we define

$$\mathcal{F}_{I\to A} = \{ X \subseteq [n] \setminus I \mid X \cup A \in \mathcal{F} \},$$

and also define its measure $\mu(\mathcal{F}_{I\to A})$ appropriately. We say a family $\mathcal{F}$ is (d, ε)-global if for any $|I| \leqslant d$ and $A \subseteq I$ it holds that $\mu(\mathcal{F}_{I\to A}) \leqslant \varepsilon$.

Theorem 7.20. *There exists $C > 0$, such that the following holds for all $d, n \in \mathbb{N}$ such that $n \geqslant 2^{C\cdot d^4}$. Let $\mathcal{F} \subseteq \binom{[n]}{k}$, and suppose that $\mathcal{F}$ is $(d, 2^{-C\cdot d^4})$-global. Then $\mu(\mathcal{F}\uparrow) \geqslant \left(1 + \frac{d}{64k}\right)\mu(\mathcal{F})$.*

Proof. Let $f = 1_{\mathcal{F}}$, $g = 1_{\mathcal{F}\uparrow}$, and consider the operator $M\colon \binom{[n]}{k} \to \binom{[n]}{k+1}$ that from a set $A \subseteq [n]$ of size k moves to a random set of size $k+1$ containing it. We also consider M as an operator $M\colon L_2\left(\binom{[n]}{k}\right) \to L_2\left(\binom{[n]}{k+1}\right)$ defined as $Mf(B) = \mathbb{E}_{A\subseteq B}[f(A)]$ (this operator is sometimes known as the *raising* or *up* operator). Note that for all $B \in \binom{[n]}{k+1}$, it holds that $g(B)Mf(B) = Mf(B)$, and that the average of Mf is the same as the average of f, i.e. $\mu(\mathcal{F})$. Thus,

$$\mu(\mathcal{F})^2 = \langle g, Mf\rangle^2 \leqslant \|g\|_2^2 \|Mf\|_2^2 = \|g\|_2^2 \langle f, M^*Mf\rangle.$$

Using the fact that the 2-norm of g squared is the measure of $\mathcal{F}\uparrow$ and rearranging, we get that

$$\mu(\mathcal{F}\uparrow) \geqslant \frac{\mu(\mathcal{F})^2}{\langle f, M^*Mf\rangle} = \frac{\mu(\mathcal{F})^2}{\Pr_{\substack{x\in_R\binom{[n]}{k} \\ y\sim MM^*x}}[x \in \mathcal{F}, y \in \mathcal{F}]}. \tag{18}$$

We next lower bound $\Pr_{\substack{x\in_R\binom{[n]}{k} \\ y\sim MM^*x}}[x \in \mathcal{F}, y \notin \mathcal{F}]$, which will give us an upper bound on the denominator. Towards this end, we relate this probability to the total influence of $1_{\mathcal{F}}$ as defined in Section 7.4.3. Note that the distribution of y conditioned on x is: with probability $1/(k+1)$ we have $y = x$, and otherwise $y = x^{(i,j)}$, where i, j are random coordinates such that $x_i \neq x_j$. Consider $z \sim \mathrm{T}x$, where T is the operator of applying a random transposition; the probability that it interchanges two coordinates i, j such that $x_i \neq x_j$ is $k(n-k)/\binom{n}{2}$, and so we get

$$\Pr_{\substack{x \in_R \binom{[n]}{k} \\ y \sim MM^* x}} [x \in \mathcal{F}, y \notin \mathcal{F}] = \frac{k}{k+1} \frac{n(n-1)}{2k(n-k)} \Pr_{\substack{x \in_R \binom{[n]}{k} \\ y \sim \mathrm{T}x}} [x \in \mathcal{F}, y \notin \mathcal{F}]$$

$$= \frac{k}{k+1} \frac{n(n-1)}{2k(n-k)} \frac{1}{2} \Pr_{\substack{x \in_R \binom{[n]}{k} \\ y \sim \mathrm{T}x}} [1_{\mathcal{F}}(x) \neq 1_{\mathcal{F}}(y)] = \frac{k}{k+1} \frac{n(n-1)}{2k(n-k)} \frac{1}{2n} I[1_{\mathcal{F}}] \geqslant \frac{1}{8k} I[1_{\mathcal{F}}],$$

which is at least $\frac{d}{64k}\mu(f)$ by Theorem 7.18 (and the fact that $\mathsf{var}(f) = \mu(f)(1-\mu(f)) \geqslant \mu(f)/2$). It follows that the denominator in (18) is at most $\mu(f)\left(1 - \frac{d}{64k}\right)$, and plugging this into (18) we get that

$$\mu(\mathcal{F}\uparrow) \geqslant \left(1 + \frac{d}{64k}\right) \mu(\mathcal{F}). \qquad \square$$

We finish this section by noting that Theorem 7.20 indeed improves on Theorem 7.19 in some range of parameters. Namely, in the case that $x = \Theta(k)$, $x \leqslant k-2$ and $n \geqslant 2^{C \cdot k^3}$. Normalizing the inequality in Theorem 7.19, we get that

$$\mu(\mathcal{F}\uparrow) \geqslant \frac{\binom{n}{k}}{\binom{n}{k+1}} \frac{\binom{n}{x+1}}{\binom{n}{x}} \mu(\mathcal{F}) = \frac{k+1}{n-k} \frac{n-x}{x+1} = \left(1 + \Theta\left(\frac{k-x}{k}\right)\right) \mu(\mathcal{F}),$$

so it is enough to note that $\mathcal{F}$ is $(d, 2^{-C \cdot d^4})$-global for $d = \left\lfloor \frac{k-x}{2} \right\rfloor$. Indeed, if $|I| = d$ and $A \subseteq I$, then

$$\mu(\mathcal{F}_{\mathcal{I} \to \mathcal{A}}) \leqslant \frac{\binom{n}{x}}{\binom{n-d}{k-|A|}} \leqslant \frac{\binom{n}{x}}{\binom{n-d}{k-d}} = \frac{n(n-1)\cdots(n-x+1)}{(n-d)(n-d-1)\cdots(n-k+1)} \frac{(k-d)!}{x!} \leqslant k^{d+1} \frac{n^x}{n^{k-d}} \leqslant k^{d+1} n^{-d},$$

which at most $2^{-C \cdot d^4}$ provided that n is large enough.

8 Proof of the level-d inequality

The goal of this section is to prove Theorem 1.6.

8.1 Proof overview

Proof overview in an idealized setting. We first describe the proof idea in an idealized setting in which derivative operators, and truncations, interact well. By that, we mean that if D is an order ℓ derivative, and f is a function, then $\mathrm{D}(f^{\leqslant d}) = (\mathrm{D}f)^{\leqslant d-\ell}$. We remark that this property holds in product spaces, but may fail in non-product domains such as S_n.

Adapting the proof of the level-d inequality from the hypercube (using Theorem 1.4 instead of standard hypercontractivity), one may easily establish a weaker version of Theorem 1.6, wherein ε^2 is replaced by $\varepsilon^{3/2}$, as follows. Take $q = \log(1/\varepsilon)$, then

$$\left\| f^{\leqslant d} \right\|_2^2 = \langle f^{\leqslant d}, f \rangle \leqslant \left\| f^{\leqslant d} \right\|_q \|f\|_{1+1/(q-1)}.$$

Since f is integer-valued, we have that $\|f\|_{1+1/(q-1)}$ is at most $\|f\|_2^{2(q-1)/q} \leqslant \varepsilon^{2(q-1)/q}$. Using the assumption of our idealized setting and Parseval, we get that for every derivative D of order ℓ we have that $\left\|\mathrm{D}(f^{\leqslant d})\right\|_2 = \left\|(\mathrm{D}f)^{\leqslant d-\ell}\right\|_2 \leqslant \|\mathrm{D}f\|_2$. Thus, using the globalness of f and both items of Claim 4.2, we get that $f^{\leqslant d}$ is $(d, 2^{O(d)}\varepsilon)$-global, and so by Theorem 1.4 we get that $\left\|f^{\leqslant d}\right\|_q \leqslant (2q)^{O(d^3)}\varepsilon$. All in all, we get that $\left\|f^{\leqslant d}\right\|_2^2 \leqslant (2q)^{O(d^3)}\varepsilon^3$, which falls short of Theorem 1.6 by a factor of ε.

The quantitative deficiency in this argument stems from the fact that $f^{\leqslant d}$ in fact is much more global than what the simplistic argument above establishes, and to show that we prove things by induction on d. This induction is also the reason we have strengthened Theorem 1.6 from the introduction to the statement above.

Returning to the real setting. To lift the assumption of the ideal setting, we return to discuss restrictions (as opposed to derivatives). Again, we would have been in good shape if restrictions were to commute with degree truncations, but this again fails, just like derivatives. Instead, we use the following observation (Claim 8.4). Suppose $k \geqslant d + \ell + 2$, and let g be a function of pure degree k, and S be a restriction of size at most ℓ. Then the restricted function g_S is perpendicular to degree $k - \ell - 1 > d$ functions, and so $(g_S)^{\leqslant d} = ((g^{\leqslant k})_S)^{\leqslant d}$.

Note that for $k = d$, this statement exactly corresponds to truncations and restrictions commuting, but the conditions of the statement always require that $k > d$ at the very least. In fact, in our setting we will have $\ell = 2d$, so we would need to use the statement with $k = 3d + 2$. Thus, to use this statement effectively we cannot apply it on our original function f, and instead have to find an appropriate choice of g such that $g^{\leqslant k}, g^{\leqslant d} \approx f^{\leqslant d}$, and moreover that they remain close under restrictions (so in particular we preserve our globalness). Indeed, we are able to design such g by applying appropriate sparse linear combinations of powers of the natural transposition operator of S_n on f.

8.2 Constructing the auxiliary function g

In this section we construct the function g.

Lemma 8.1. *There is an absolute constant $C > 0$, such that the following holds. Suppose $n \geqslant 2^{C\cdot d^3}$, and let T be the adjacency operator of the transpositions graph (see Section 7.3). There exists a polynomial P with $\|P\| \leqslant 2^{C\cdot d^4}$ such that*

$$\left\|P(\mathrm{T})(f^{\leqslant 4d}) - f^{\leqslant d}\right\|_2 \leqslant \left(\frac{1}{n}\right)^{19d} \left\|f^{\leqslant 4d}\right\|_2.$$

Proof. Let

$$Q(z) = \sum_{i=1}^{d} \prod_{j\in[4d]\setminus\{i\}} \left(\frac{z^n - e^{-2j}}{e^{-2i} - e^{-2j}}\right)^{20d},$$

and define $P(z) = 1 - (1 - Q(z))^{20d}$. We first prove the upper bound on $\|P\|$; note that

$$\|Q\| \leqslant \sum_{i=1}^{d} \prod_{j\in[4d]\setminus\{i\}} \left\|\frac{z^n - e^{-2j}}{e^{-2i} - e^{-2j}}\right\|^{20d} = \sum_{i=1}^{d} \prod_{j\in[4d]\setminus\{i\}} \left(\frac{1 + e^{-2j}}{e^{-2i} - e^{-2j}}\right)^{20d} = 2^{O\left(d^3\right)},$$

so $\|P\| \leqslant (1 + 2^{O(d^3)})^{20d} = 2^{O(d^4)}$.

Next, we show that for $g = P(\mathrm{T})f$, it holds that $\left\|g^{\leqslant 4d} - f^{\leqslant d}\right\|_2 \leqslant \left(\frac{1}{n}\right)^{10d} \left\|f^{\leqslant 4d}\right\|_2$, and we do so by eigenvalue considerations. Let $d < \ell \leqslant 4d$, and let λ be an eigenvalue of T corresponding to a function of pure degree ℓ. Since $\ell \leqslant n/2$, Lemma 7.10 implies that $\lambda = 1 - \frac{2\ell}{n} + O\left(\frac{\ell^2}{n^2}\right)$, and so $\lambda^n = e^{-2\ell} \pm O\left(\frac{d^2}{n}\right)$. Thus, as each one of the products in $Q(\lambda)$ contains a term for ℓ, we get that

$$|Q(\lambda)| \leqslant d \cdot \left(\frac{2^{O(d^2)}}{n}\right)^{20d} \leqslant \frac{2^{O(d^3)}}{n^{20d}},$$

so $|P(\lambda)| = 1 - (1 - \frac{2^{O(d^3)}}{n^{20d}})^d \leqslant \frac{1}{n^{19d}}$. Next, let $\ell \leqslant d$, and let λ be an eigenvalue of T corresponding to a function of pure degree ℓ. As before, $\lambda^n = e^{-2\ell} \pm O\left(\frac{d^2}{n}\right)$, but now in $Q(\lambda)$ there is one product that omits the term for ℓ. A direct computation gives that

$$Q(\lambda) = \prod_{j \in [4d] \setminus \{\ell\}} \left(\frac{\lambda^n - e^{-2j}}{e^{-2\ell} - e^{-2j}}\right)^{20d} + \frac{2^{O(d^3)}}{n^{20d}} = \prod_{j \in [4d] \setminus \{\ell\}} \left(1 - O\left(\frac{2^{O(d)}}{n}\right)\right) + \frac{2^{O(d^3)}}{n^{20d}},$$

so $Q(\lambda) = 1 - O\left(\frac{2^{O(d)}}{n}\right)$. Thus,

$$|P(\lambda) - 1| = O\left(\frac{2^{O(d^2)}}{n^{20d}}\right) \leqslant \frac{1}{n^{19d}}.$$

It follows that $g^{\leqslant 4d} - f^{\leqslant d} = \sum_{\ell=0}^{4d} c_\ell f^{=\ell}$ for $|c_\ell| \leqslant \frac{1}{n^{19d}}$, and the result follows from Parseval. $\square$

8.3 Properties of Cayley operators and restrictions

In this section, we study random walks along Cayley graphs on S_n. The specific transition operator we will later be concerned with is the transposition operator from Lemma 7.10 and its powers, but we will present things in greater generality.

8.3.1 Random walks

Definition 8.2. *A Markov chain* M *on* S_n *is called a Cayley random walk if for any* $\sigma, \tau, \pi \in S_n$*, the transition probability from* σ *to* τ *is the same as the transition probability from* $\sigma\pi$ *to* $\tau\pi$*.*

In other words, a Markov chain M is called Cayley if the transition probability from σ to τ is only a function of $\sigma\tau^{-1}$. We will be interested in the interaction between random walks and restrictions, and towards this end we first establish the following claim, asserting that a Cayley random walk either never transitions between two restrictions T and T', or can always transition between the two.

Claim 8.3. *Suppose* M *is a Cayley random walk on* S_n*, let* $i_1, \ldots, i_t \in [n]$ *be distinct, and let* $T = \{(i_1, j_1), \ldots, (i_t, j_t)\}$*,* $T' = \{(i_1, j'_1), \ldots, (i_t, j'_t)\}$ *be consistent sets. Then one of the following two must hold:*

1. $\Pr_{\substack{u \in S_n^{T'} \\ v \sim \mathrm{M}v}} \left[v \in S_n^T\right] = 0.$

2. *For all* $\pi \in S_n^T$, *it holds that* $\Pr_{\substack{u \in S_n^{T'} \\ v \sim \mathrm{M}v}} \left[v = \pi \right] > 0$.

Proof. If the first item holds then we're done, so let us assume otherwise. Then there are $u \in S_n^{T'}$, $v \in S_n^T$ such that M has positive probability of transitioning from u to v. Denoting $\tau = uv^{-1}$, we note that $\tau(j_\ell) = j'_\ell$ for all $\ell = 1, \ldots, t$. Fix $\pi \in S_n^T$. Since M is a Cayley operator, the transition probability from $\tau\pi$ to π is positive, and since $\tau\pi$ is in $S_n^{T'}$, the proof is concluded. $\square$

If M satisfies the second item of the above claim with T and T', we say that M is compatible with (T, T').

8.3.2 Degree decomposition on restrictions

Let $T = \{(i_1, j_1), \ldots, (i_t, j_t)\}$ be consistent. A function $f \in L^2(S_n^T)$ is called a d-junta, if there is $S \subseteq [n] \setminus \{i_1, \ldots, i_t\}$ of size d such that $f(\pi)$ only depends on $\pi(i)$ for $i \in S$ (we say that $f(\pi)$ only depends on $\pi(S)$). With this definition in hand, we may define the space of degree d functions on S_n^T, denoted by $V_d(S_n^T)$, as the span of all d-juntas, and subsequently define projections onto this subspaces. That is, for each $f \in L^2(S_n^T)$ we denote by $f^{\leqslant d}$ the projection of f onto $V_d(S_n^T)$. Finally, we define the pure degree d part of f as $f^{=d} = f^{\leqslant d} - f^{\leqslant d-1}$.

We have the following basic property of pure degree d functions.

Claim 8.4. *Suppose that* $f\colon S_n \to \mathbb{R}$ *is of pure degree* d*. Let* T *be a set of size* $\ell < d$*. Then* f_T *is orthogonal to all functions in* $V_{d-1-\ell}$*.*

Proof. Clearly, it is enough to show that f_T is orthogonal to all $(d-1-\ell)$-juntas. Fix $g\colon S_n^T \to \mathbb{R}$ to be a $(d-1-\ell)$-junta, and let h be its extension to S_n by setting it to be 0 outside S_n^T. Then h is a $(d-1)$-junta, and so

$$0 = \langle f, h \rangle = \frac{(n-\ell)!}{n!} \langle f_T, g \rangle .$$

$\square$

8.3.3 Extension to functions

Any random walk M on S_n extends to an operator on functions on S_n, which maps $f\colon S_n \to \mathbb{R}$ to the function $\mathrm{M}f\colon S_n \to \mathbb{R}$ given by

$$\mathrm{M}f(\pi) = \mathop{\mathbb{E}}_{\substack{u \in S_n \\ v \sim \mathrm{M}u}} [f(u) \mid v = \pi].$$

8.4 Strengthening Proposition 3.1

Our main goal in this section is to prove the following statement that both strengtheners and generalizes Proposition 3.1.

Proposition 8.5. *Let* $f\colon S_n \to \mathbb{R}$*. Let* M *be a Cayley random walk on* S_n*, let* $g = \mathrm{M}f$*, and let* $T = \{(i_1, j_1), \ldots, (i_t, j_t)\}$ *be a consistent set. Then for all* d*,*

$$\| (g_T)^{\leqslant d} \|_2 \leqslant \max_{\substack{T' = \{(i_1, j'_1), \ldots, (i_t, j'_t)\} \\ \mathrm{M} \text{ compatible with } (T, T')}} \| (f_{T'})^{\leqslant d} \|_2.$$

Let M be a Cayley random walk and let $T = \{(i_1, j_1), \ldots, (i_t, j_t)\}$ and $T' = \{(i_1, j_1'), \ldots, (i_t, j_t')\}$ be consistent so that M is compatible with (T, T'). Put $I = \{(i_1, i_1), \ldots, (i_t, i_t)\}$. Define the operator $\mathrm{M}_{S_n^T \to S_n^{T'}} \colon L^2(S_n^T) \to L^2(S_n^{T'})$ in the following way: given a function $f \in L^2(S_n^T)$, we define

$$\mathrm{M}_{S_n^{T'} \to S_n^T} f(\pi) = \mathop{\mathbb{E}}_{\substack{u \in_R S_n^T \\ v \sim \mathrm{M}u}} \left[f(u) \mid v = \pi \right].$$

Drawing inspiration from the proof of Proposition 3.1, we study the operator $\mathrm{M}_{S_n^T \to S_n^{T'}}$. Since we are also dealing with degree truncations, we have to study its interaction with this operator. Indeed, a key step in the proof is to show that the two operators commute, in the following sense: for all $d \in \mathbb{N}$ and $f \in L^2(S_n^T)$, it holds that

$$\left(\mathrm{M}_{S_n^T \to S_n^{T'}} f\right)^{=d} = \mathrm{M}_{S_n^T \to S_n^{T'}} \left(f^{=d}\right).$$

Towards this end, we view $L^2(S_n^T)$ (and similarly $L^2(S_n^{T'})$) as a right S_n^I-module using the following operation: a function-permutation pair $(f, \pi) \in L^2(S_n^T) \times S_n^I$ is mapped to a function $f^\pi \in L^2(S_n^T)$ defined as

$$f^\pi(\sigma) = f(\sigma \pi^{-1}).$$

Claim 8.6. *With the setup above,* $\mathrm{M}_{S_n^T \to S_n^{T'}} \colon L^2\left(S_n^T\right) \to L^2\left(S_n^{T'}\right)$ *is a homomorphism of* S_n^I*-modules.*

Proof. The proof is essentially the same as the proof of Lemma 5.1, and is therefore omitted. $\square$

Therefore, it is sufficient to prove that any homomorphism commutes with taking pure degree d part, which is the content of the following claim.

Claim 8.7. *Let* T, T' *be consistent as above, and let* $\mathrm{A} \colon L^2(S_n^T) \to L^2(S_n^{T'})$ *be a homomorphism of right* S_n^I*-modules. Then for all* $f \in L^2(S_n^T)$ *we have that*

$$(\mathrm{A}f)^{=d} = \mathrm{A}\left(f^{=d}\right).$$

Proof. We first claim that A preserves degrees, i.e. $\mathrm{A}V_d(S_n^T) \subseteq V_d(S_n^{T'})$. To show this, it is enough to note that if $f \in L^2(S_n^T)$ is a d-junta, then $\mathrm{A}f$ is a d-junta. Let f be a d-junta, and suppose that $S \subseteq [n]$ is a set of size at most d such that $f(\sigma)$ only depends on $\sigma(S)$. Then for any π that has S as fixed points, we have that $f(\sigma) = f(\sigma\pi^{-1}) = f^\pi(\sigma)$, so $f = f^\pi$. Applying A and using the previous claim we get that $\mathrm{A}f = \mathrm{A}f^\pi = (\mathrm{A}f)^\pi$. This implies that $\mathrm{A}f$ is invariant under any permutation that keeps S as fixed points, so it is an S-junta.

Let $V_{=d}(S_n^T)$ be the space of functions of pure degree d, i.e. $V_d(S_n^T) \cap V_{d-1}(S_n^T)^\perp$. We claim that A also preserves pure degrees, i.e. $\mathrm{A}V_{=d}(S_n^T) \subseteq V_{=d}(S_n^{T'})$. By the previous paragraph it is enough to show that if $f \in V_{=d}(S_n^T)$, then $\mathrm{A}f$ is orthogonal to $V_{d-1}(S_n^{T'})$. Letting A^* be the adjoint operator of A, it is easily seen that $\mathrm{A}^* \colon L^2(S_n^{T'}) \to L^2(S_n^T)$ is also a homomorphism between right S_n^I-modules, and by the previous paragraph it follows that A^* preserves degrees. Thus, for any $g \in V_{d-1}(S_n^{T'})$ we have that $\mathrm{A}^* g \in V_{d-1}(S_n^T)$, and so

$$\langle \mathrm{A}f, g \rangle = \langle f, \mathrm{A}^* g \rangle = 0.$$

We can now prove the statement of the claim. Fix $f \in L^2(S_n^T)$ and d. Then by the above paragraph, $\mathrm{A}\left(f^{=d}\right) \in V_{=d}(S_n^{T'})$, and by linearity of A we have $\sum_d \mathrm{A}\left(f^{=d}\right) = \mathrm{A}f$. The claim follows from the uniqueness of the degree decomposition. $\square$

We define a transition operator on restrictions as follows. From a restriction $T = \{(i_1, j_1), \dots, (i_t, j_t)\}$, we sample $T' \sim N(T)$ as follows. Take $\pi \in S_n^T$ uniformly, sample $\sigma \sim \mathrm{M}\pi$, and then let T' be $\{(i_1, \sigma(i_1)), \dots, (i_t, \sigma(i_t))\}$. The following claim is immediate:

Claim 8.8. $(\mathrm{M}f)_T = \mathbb{E}_{T' \sim N(T)}\left[\mathrm{M}_{S_n^{T'} \to S_n^T} f_{T'}\right]$.

We are now ready to prove Proposition 8.5.

Proof of Proposition 8.5. By Claim 8.8, we have $g_T = \mathbb{E}_{T' \sim T}\mathrm{M}_{S_n^{T'} \to S_n^T} f_{T'}$. Using Claim 8.7 and the linearity of the operator $f \mapsto f^{=d}$, we get

$$(g_T)^{=d} = \mathbb{E}_{T' \sim N(T)}\mathrm{M}_{S_n^{T'} \to S_n^T}\left((f_{T'})^{=d}\right).$$

Summing this up using linearity again, we conclude that

$$(g_T)^{\leqslant d} = \mathbb{E}_{T' \sim N(T)}\mathrm{M}_{S_n^{T'} \to S_n^T}\left((f_{T'})^{\leqslant d}\right).$$

Taking norms and using the triangle inequality gives us that

$$\| (g_T)^{\leqslant d} \|_2 \leqslant \mathbb{E}_{T' \sim N(T)}\left\|\mathrm{M}_{S_n^{T'} \to S_n^T}\left((f_{T'})^{\leqslant d}\right)\right\|_2 \leqslant \max_{T' :\ \mathrm{M}\text{ consistent with }(T,T')}\left\|\mathrm{M}_{S_n^{T'} \to S_n^T}\left((f_{T'})^{\leqslant d}\right)\right\|_2.$$

The proof is now concluded by appealing to Fact 3.2. □

8.5 A weak level-d inequality

The last ingredient we will need in the proof of Theorem 1.6 is a weak version of the level-d inequality, which does not take the globalness of f into consideration.

Lemma 8.9. *Let C be sufficiently large, let $n \geqslant \log(1/\varepsilon)^d C^{d^2}$, and let $f \colon S_n \to \{0, 1\}$ satisfy $\|f\|_2 \leqslant \varepsilon$. Then*

$$\|f^{\leqslant d}\|_2 \leqslant n^d \log(1/\varepsilon)^{O(d)} \varepsilon^2.$$

Proof. Set $q = \log(1/\varepsilon)$, and without loss of generality assume q is an even integer (otherwise we may change q by a constant factor to ensure that). Using Hölder's inequality, Lemma 5.14, and the fact that $\|f\|_{q/(q-1)} = O\left(\varepsilon^2\right)$, we obtain

$$\left\|f^{\leqslant d}\right\|_2^2 = \left\langle f^{\leqslant d}, f\right\rangle \leqslant \left\|f^{\leqslant d}\right\|_q \|f\|_{q/(q-1)} \leqslant \log(1/\varepsilon)^{O(d)} n^d \left\|f^{\leqslant d}\right\|_2 \varepsilon^2,$$

and the lemma follows by rearranging. □

8.6 Interchanging truncations and derivatives with small errors

Lemma 8.10. *There is $C > 0$, such that the following holds for $n \geqslant 2^{C \cdot d^3}$. For all derivatives D of order $t \leqslant d$ we have:*

$$\left\|D\left(f^{\leqslant d}\right)\right\|_2 \leqslant 2^{O(d)^4} \max_{t-\text{derivative } \mathrm{D}'}\left\|(\mathrm{D}'f)^{\leqslant d-t}\right\|_2 + \left(\frac{1}{n}\right)^{10d}\left\|f^{\leqslant 4d}\right\|_2.$$

Proof. Let $\mathrm{T}_d = P(\mathrm{T})$ be as in Lemma 8.1, and write $f^{\leqslant d} = \mathrm{T}_d(f^{\leqslant 4d}) + g$, where $\|g\|_2 \leqslant n^{-19d} \left\|f^{\leqslant 4d}\right\|_2$. Let S be a consistent restriction of t coordinates, and let D be a derivative along S. Then there is $R \subseteq L$ of size t such that $\mathrm{D}f = (\mathrm{L}f)_{S\to R}$. By Claim 4.3, the degree of $\mathrm{D}(f^{\leqslant d})$ is at most $d-t$, thus

$$\mathrm{D}(f^{\leqslant d}) = \left(\mathrm{D}(f^{\leqslant d})\right)^{\leqslant d-t}. \tag{19}$$

We want to compare the right-hand side with $\left(\mathrm{D}\left(\mathrm{T}_d f\right)\right)^{\leqslant d-t}$, but first we show that in it one may truncate all degrees higher than $4d$ in f. Note that by Claim 8.7, for each $k > 4d$ the function $\mathrm{T}_d f^{=k}$ has pure degree k, so $\mathrm{D}(\mathrm{T}_d f^{=k})$ is perpendicular to degree $k-t-1$ functions. Since $k - 2t - 1 \geqslant d - t$, we have that its level $d-t$ projection is 0, so $\left(\mathrm{D}\left(\mathrm{T}_d f\right)\right)^{\leqslant d-t} = \left(\mathrm{D}\left(\mathrm{T}_d f^{\leqslant 4d}\right)\right)^{\leqslant d-t}$. It follows that

$$\begin{aligned}\left\|\mathrm{D}(f^{\leqslant d}) - \left(\mathrm{D}\left(\mathrm{T}_d f\right)\right)^{\leqslant d-t}\right\|_2 = \left\|\left(\mathrm{D}\left(f^{\leqslant d} - \mathrm{T}_d(f^{\leqslant 4d})\right)\right)^{\leqslant d-t}\right\|_2 &\leqslant \|\mathrm{D}g\|_2 \leqslant n^{2t}\|g\|_2 \\ &\leqslant n^{2t-19d}\left\|f^{\leqslant 4d}\right\|_2. \end{aligned}\tag{20}$$

Our task now is to bound $\left\|\left(\mathrm{D}\left(\mathrm{T}_d f\right)\right)^{\leqslant d-t}\right\|_2$. Since T commutes with Laplacians, it follows that T_d also commutes with Laplacians, and so

$$\left(\mathrm{D}\left(\mathrm{T}_d f\right)\right)^{\leqslant d-t} = \left((\mathrm{L}\mathrm{T}_d f)_{S\to R}\right)^{\leqslant d-t} = \left((\mathrm{T}_d \mathrm{L} f)_{S\to R}\right)^{\leqslant d-t}. \tag{21}$$

By Proposition 8.5, for all i and $h\colon S_n \to \mathbb{R}$ we have

$$\|\left(\left(\mathrm{T}^i h\right)_S\right)^{\leqslant d}\|_2 \leqslant \max_{S'=\left\{\left(i_1,j_1'\right),\ldots,\left(i_t,j_t'\right)\right\}} \left\|\left(h_{S'}\right)^{\leqslant d}\right\|_2,$$

and so

$$\|\left(\left(\mathrm{T}_d h\right)_S\right)^{\leqslant d}\|_2 \leqslant \|P\| \max_{S'=\left\{\left(i_1,j_1'\right),\ldots,\left(i_t,j_t'\right)\right\}} \left\|\left(h_{S'}\right)^{\leqslant d}\right\|_2 \leqslant 2^{O(d^4)} \max_{S'=\left\{\left(i_1,j_1'\right),\ldots,\left(i_t,j_t'\right)\right\}} \left\|\left(h_{S'}\right)^{\leqslant d}\right\|_2.$$

Applying this for $h = Lf$ gives that

$$\left\|\left((\mathrm{T}_d \mathrm{L} f)_{S\to R}\right)^{\leqslant d-t}\right\|_2 \leqslant 2^{O\left(d^4\right)} \max_{R'} \left\|\left((\mathrm{L}f)_{S\to R'}\right)^{\leqslant d-t}\right\|_2 = 2^{O\left(d^4\right)} \max_{D'} \left\|\left(\mathrm{D}'f\right)^{\leqslant d-t}\right\|_2, \tag{22}$$

where the last transition is by the definition of derivatives. Combining (20), (21), (22) and using the triangle inequality finishes the proof. □

8.7 Proof of the level-d inequality

We end this section by deriving the following proposition, which by Claim 4.2 implies Theorem 1.6.

Proposition 8.11. *There exists an absolute constant $C > 0$ such that the following holds for all $d \in \mathbb{N}$, $\varepsilon > 0$ and $n \geqslant 2^{C\cdot d^3}\log(1/\varepsilon)^{C\cdot d}$. Let $f\colon S_n \to \mathbb{Z}$ be a function, such that for all $t \leqslant d$ and all t-derivatives D we have $\|\mathrm{D}f\|_2 \leqslant \varepsilon$. Then*

$$\left\|f^{\leqslant d}\right\|_2 \leqslant 2^{Cd^4}\varepsilon^2 \log\left(1/\varepsilon\right)^{Cd}.$$

Proof. The proof is by induction on d. If $d=0$, then

$$\left\|f^{\leqslant d}\right\|_2 = \left|\mathbb{E}\left[f(\pi)\right]\right| \leqslant \mathbb{E}\left[|f(\pi)|^2\right] = \|f\|_2^2 \leqslant \varepsilon^2,$$

where in the second transition we used the fact that f is integer-valued.

We now prove the inductive step. Fix $d \geqslant 1$. Let $1 \leqslant t \leqslant d$, and let D be a t-derivative. By Lemma 8.10, there is an absolute constant $C_1 > 0$ such that

$$\left\|\mathrm{D}\left(f^{\leqslant d}\right)\right\|_2 \leqslant e^{C_1\left(d^4\right)} \max_{\mathrm{D}' \text{ a } t-\text{derivative}} \left\|\left(\mathrm{D}'f\right)^{\leqslant d-t}\right\|_2 + n^{-10d}\left\|f^{\leqslant 4d}\right\|_2. \tag{23}$$

Fix D'. The function $\mathrm{D}'f$ takes integer values and is defined on a domain that is isomorphic to S_{n-t}, so by the induction hypothesis we have

$$\left\|\left(\mathrm{D}'f\right)^{\leqslant d-t}\right\|_2 \leqslant e^{C(d-t)^4}\varepsilon^2 \log\left(\frac{1}{\varepsilon}\right)^{C(d-t)}.$$

As for $\|f^{\leqslant 4d}\|_2^2$, applying Lemma 8.9 we see it is at most $n^{8d}\varepsilon^4\log^{Cd}(1/\varepsilon)$. Plugging these two estimates into (23) we get that

$$\left\|\mathrm{D}\left(f^{\leqslant d}\right)\right\|_2 \leqslant e^{Cd^4}\varepsilon^2\log^{C\cdot d}(1/\varepsilon),$$

provided that C is sufficiently large.

If

$$\left\|f^{\leqslant d}\right\|_2^2 \leqslant e^{2Cd^4}\varepsilon^4\log\left(\frac{1}{\varepsilon}\right)^{2Cd}$$

we're done, so assume otherwise. We get that $\left\|\mathrm{D}'\left(f^{\leqslant d}\right)\right\|_2 \leqslant \|f^{\leqslant d}\|_2$ for all derivatives of order at most d, and from Claim 4.3, $\left\|\mathrm{D}'\left(f^{\leqslant d}\right)\right\|_2 = 0$ for higher-order derivatives, and so by Claim 4.2, the function $f^{\leqslant d}$ is $(2d, 4^d\|f^{\leqslant d}\|_2)$-global, and by Lemma 3.5, we get that $f^{\leqslant d}$ is $4^d\|f^{\leqslant d}\|_2$-global with constant 4^8. In this case, we apply the standard argument as presented in the overview, as outlined below.

Set $q = \log(1/\varepsilon)$, and without loss of generality assume q is an even integer (otherwise we may change q by a constant factor to ensure that). Set $\rho = \frac{1}{(10q4^8)^2}$. From Lemmas 5.1, 5.4 we have that $\mathrm{T}^{(\rho)}$ preserves degrees, and so by Corollary 5.11 we get

$$\left\|f^{\leqslant d}\right\|_2^2 \leqslant \rho^{-C_2\cdot d}\langle \mathrm{T}^{(\rho)}f^{\leqslant d}, f^{\leqslant d}\rangle = \rho^{-C_2\cdot d}\langle \mathrm{T}^{(\rho)}f^{\leqslant d}, f\rangle \leqslant \rho^{-C_2\cdot d}\left\|\mathrm{T}^{(\rho)}f^{\leqslant d}\right\|_q \|f\|_{q/(q-1)},$$

where we also used Hölder's inequality. By Theorem 3.3, we have $\left\|\mathrm{T}^{(\rho)}f^{\leqslant d}\right\|_q \leqslant 4^d\left\|f^{\leqslant d}\right\|_2$, and by a direction computation $\|f\|_{q/(q-1)} \leqslant \varepsilon^{2(q-1)/q}$. Plugging these two estimates into the inequality above and rearranging yields that

$$\left\|f^{\leqslant d}\right\|_2^2 \leqslant \rho^{-2C_2\cdot d}4^{2d}\|f\|_{q/(q-1)}^2 \leqslant \rho^{-3C_2\cdot d}\varepsilon^4 = 2^{6C_2\log(10C)}\varepsilon^4\log^{6C_2\cdot d}(1/\varepsilon) \leqslant 2^{C\cdot d^4}\varepsilon^4\log^{C\cdot d}(1/\varepsilon),$$

for large enough C. □

8.8 Deducing the strong level-d inequality: proof of Theorem 1.7

Let $\delta = 2^{C_1 \cdot d^4} \varepsilon^2 \log^{C_1 \cdot d}(1/\varepsilon)$ for sufficiently large absolute constant C_1. By Claim A.1 we get that $f^{\leqslant d}$ is δ-global with constant 4^8. Set $q = \log(1/\|f\|_2)$, and let $\rho = 1/(10 \cdot 4^8 \cdot q)^2$ be from Theorem 3.3. From Lemmas 5.1, 5.4 we have that $\mathrm{T}^{(\rho)}$ preserves degrees, and so by Corollary 5.11 we get

$$\left\|f^{\leqslant d}\right\|_2^2 = \langle f^{\leqslant d}, f^{\leqslant d}\rangle \leqslant \rho^{-O(d)} \langle f^{\leqslant d}, \mathrm{T}^{(\rho)} f^{\leqslant d}\rangle = \rho^{-O(d)} \langle f, \mathrm{T}^{(\rho)} f^{\leqslant d}\rangle \leqslant \rho^{-O(d)} \|f\|_{q/(q-1)} \left\|\mathrm{T}^{(\rho)} f^{\leqslant d}\right\|_q .$$

Using $\|f\|_{q/(q-1)} \leqslant \|f\|_2^{2(q-1)/q} = \|f\|_2^2 \|f\|_2^{-2/q} \leqslant O(\|f\|_2^2)$ and Theorem 3.3 to bound $\left\|\mathrm{T}^{(\rho)} f^{\leqslant d}\right\|_q \leqslant \delta$, it follows that

$$\left\|f^{\leqslant d}\right\|_2^2 \leqslant \rho^{-O(d)} \|f\|_2^2 \delta \leqslant 2^{C \cdot d^4} \|f\|_2^2 \varepsilon^2 \log^{C \cdot d}(1/\varepsilon),$$

where we used $\|f\|_2^2 \leqslant \varepsilon$. □

References

[1] E. Bannai and T. Ito. *Algebraic combinatorics. I.* The Benjamin/Cummings Publishing Co., Inc., Menlo Park, CA, 1984. Association schemes.

[2] J. Bourgain. On the distribution of the Fourier spectrum of Boolean functions. *Israel J. of Math.*, (131):269–276, 2002.

[3] M. Braverman, S. Khot, and D. Minzer. On rich \$2\$-to-\$1\$ games. *Electronic Colloquium on Computational Complexity (ECCC)*, 26:141, 2019.

[4] P. Diaconis and M. Shahshahani. Time to reach stationarity in the Bernoulli–Laplace diffusion model. *SIAM Journal on Mathematical Analysis*, 18(1):208–218, 1987.

[5] C. F. Dunkl. Orthogonal functions on some permutation groups. In *Relations between combinatorics and other parts of mathematics (Proc. Sympos. Pure Math., Ohio State Univ., Columbus, Ohio, 1978)*, Proc. Sympos. Pure Math., XXXIV, pages 129–147. Amer. Math. Soc., Providence, R.I., 1979.

[6] S. Eberhard. Product mixing in the alternating group. *arXiv preprint arXiv:1512.03517*, 2015.

[7] D. Ellis, E. Friedgut, and H. Pilpel. Intersecting families of permutations. *Journal of the American Mathematical Society*, 24(3):649–682, 2011.

[8] Y. Filmus. Orthogonal basis for functions over a slice of the Boolean hypercube. *Electronic Journal of Combinatorics*, 23(1):P1.23, 2016.

[9] Y. Filmus and E. Mossel. Harmonicity and invariance on slices of the Boolean cube. *Probability Theory and Related Fields*, 175(3–4):721–782, 2019.

[10] Y. Filmus, R. O'Donnell, and X. Wu. A log-sobolev inequality for the multislice, with applications. In *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, pages 34:1–34:12, 2019.

[11] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[12] E. Friedgut and J. Bourgain. Sharp thresholds of graph properties, and the k-SAT problem. *Journal of the American mathematical Society*, 12(4):1017–1054, 1999.

[13] W. Fulton and J. Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, 2013.

[14] W. T. Gowers. Quasirandom groups. *Combinatorics, Probability and Computing*, 17(3):363–387, 2008.

[15] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *FOCS 1988*, pages 68–80, 1988.

[16] P. Keevash. Shadows and intersections: stability and new proofs. *Adv. Math.*, 218(5):1685–1703, 2008.

[17] P. Keevash, N. Lifshitz, E. Long, and D. Minzer. Hypercontractivity for global functions and sharp thresholds. *arXiv preprint arXiv:1906.05568*, 2019.

[18] P. Keevash, N. Lifshitz, E. Long, and D. Minzer. Forbidden intersections for codes. 2020.

[19] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 767–775, 2002.

[20] S. Khot, D. Minzer, and M. Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *FOCS 2018*, pages 592–601, 2018.

[21] T.-Y. Lee and H.-T. Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.*, 26(4):1855–1873, 1998.

[22] N. Lifshitz and D. Minzer. Noise sensitivity on the p -biased hypercube. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1205–1226, 2019.

[23] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *FOCS 2005*, pages 21–30, 2005.

[24] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[25] R. O'Donnell and K. Wimmer. KKL, Kruskal-Katona, and monotone nets. *SIAM J. Comput.*, 42(6):2375–2399, 2013.

[26] J. Salez. A sharp log-Sobolev inequality for the multislice. *ArXiv*, abs/2004.05833, 2020.

A Missing proofs

A.1 Globalness of f implies globalness of $f^{\leqslant d}$

Claim A.1. *There exists an absolute constant $C > 0$ such that the following holds for all $n, d \in \mathbb{N}$ and $\varepsilon > 0$ satisfying $n \geqslant 2^{C \cdot d^3} \log(1/\varepsilon)^{C \cdot d}$. Suppose $f \colon S_n \to \mathbb{Z}$ is $(2d, \varepsilon)$-global. Then for all $j \leqslant d$, the function $f^{\leqslant j}$ is*

1. $(2j, 2^{O(j^4)}\varepsilon^2 \log^{O(j)}(1/\varepsilon))$-*global.*

2. $2^{O(j^4)}\varepsilon^2 \log^{O(j)}(1/\varepsilon)$-*global with constant* 4^8.

Proof. If $j = 0$, then the claim is clear as $f^{\leqslant j}$ is just the constant $\mathbb{E}\left[f(\pi)\right]$, and its absolute value is at most $\|f\|_2^2 \leqslant \varepsilon^2$.

Suppose $j \geqslant 1$ and let D be a derivative of order $1 \leqslant r \leqslant j$, then by Claim 4.2 we have $\|\mathrm{D}f\|_2 \leqslant 2^{2j}\varepsilon$. Therefore, applying Proposition 8.11 on $\mathrm{D}f$, we get that

$$\left\|(\mathrm{D}f)^{\leqslant j-1}\right\|_2 \leqslant 2^{O((j-1)^4)}\varepsilon^2 \log^{O(j)}(1/\varepsilon).$$

Using Lemma 8.10 we get that

$$\left\|\mathrm{D}(f^{\leqslant j})\right\|_2 \leqslant 2^{O(j)^4} \max_{1-\text{derivative } \mathrm{D}'} \left\|(\mathrm{D}'f)^{\leqslant j-1}\right\|_2 + \left(\frac{1}{n}\right)^{10j} \left\|f^{\leqslant 4j}\right\|_2 \leqslant 2^{O(j^4)}\varepsilon^2 \log^{O(j)}(1/\varepsilon),$$

where in the last inequality we our earlier estimate and Lemma 8.9. For derivatives of order higher than j, we have that $\mathrm{D}(f^{\leqslant j}) = 0$ from Claim 4.3. Thus, Claim 4.2 implies that $f^{\leqslant j}$ is $(2j, 2^{O(j^4)}\varepsilon^2 \log^{O(j)}(1/\varepsilon))$-global. The second item immediately follows from Lemma 3.5. $\square$

A.2 Proof of Theorem 7.13

Our proof will make use of the following simple fact.

Fact A.2. *Let* $g\colon S_n \to \mathbb{R}$.

1. *We have the Poincaré inequality:* $\mathsf{var}(g) \leqslant \frac{1}{n} \sum\limits_{\mathrm{L}_1} \|\mathrm{L}_1 g\|_2^2$, *where the sum is over all* 1*-Laplacians.*

2. *We have* $I[g] = \frac{2}{n-1} \sum\limits_{\mathrm{L}_1} \|\mathrm{L}_1 g\|_2^2$, *where again the sum is over all* 1*-Laplacians.*

Proof. The second item is straightforward by the definitions, and we focus on the first one. Let $\tilde{\mathrm{L}}g = \mathbb{E}_{\mathrm{L}_1}\left[\mathrm{L}_1 g\right] = (I - \mathrm{T})g$. If $\alpha_{d,r}$ is an eigenvalue of T corresponding to a function from $V_{=d}(S_n)$, then by the second item in Lemma 7.10 we have $\alpha_{d,r} \leqslant 1 - \frac{d}{n-1}$.

Note that we may find an orthonormal basis of $V_{=d}(S_n)$ consisting of eigenvectors of T, and therefore we may first write $g = \sum\limits_{d} g^{=d}$ where $g^{=d} \in V_{=d}(S_n)$, and then further decompose each g_d to $g_d = \sum\limits_{r=0}^{r_d} g_{d,r}$ where $g_{d,r} \in V_{=d}(S_n)$ are all orthogonal and eigenvectors of T. We thus get

$$\langle g, \tilde{\mathrm{L}}g\rangle = \sum_{d}\sum_{r=0}^{r_d}(1-\alpha_{d,r}) \left\|g^{d,r}\right\|_2^2 \geqslant \sum_{d}\sum_{r=0}^{r_d} \frac{d}{n-1} \left\|g^{d,r}\right\|_2^2 = \sum_{d} \frac{d}{n-1} \left\|g^{=d}\right\|_2^2 \geqslant \frac{1}{n-1}\mathsf{var}(g). \quad (24)$$

On the other hand,

$$\langle g, \tilde{L}g\rangle = \mathop{\mathbb{E}}_{\pi}\left[\mathop{\mathbb{E}}_{\tau \text{ a transposition}}\left[g(\pi)(g(\pi) - g(\pi\circ\tau))\right]\right] = \frac{1}{2}\mathop{\mathbb{E}}_{\tau \text{ a transposition}}\left[\mathop{\mathbb{E}}_{\pi}\left[(g(\pi) - g(\pi\circ\tau))^2\right]\right],$$

which is the same as $\frac{1}{2\binom{n}{2}} \sum\limits_{\mathrm{L}_1} \|\mathrm{L}_1 g\|_2^2$. Combining this with the previous lower bound gives the first item. $\square$

Proof of Theorem 7.13. Let $f = 1_S$. Then $I[f] = \frac{n-1}{2}\Pr_{\substack{\pi\in S_n\\ \sigma\sim \mathrm{T}\pi}}[f(\pi)\neq f(\sigma)]$, and arithmetizing that we have that it is equal to $\frac{n-1}{2}\langle f,(I-\mathrm{T})f\rangle$. Thus, writing $f = f^{=0}+f^{=1}+\dots$, where $f^{=j}\in V_{=j}(S_n)$, we have, as in inequality (24), that

$$\frac{n-1}{2}\langle f,(I-\mathrm{T})f\rangle \geqslant \frac{n-1}{2}\sum_{j=0}^{n}\frac{j}{n-1}\left\|f^{=j}\right\|_2^2 \geqslant \frac{d}{2}\left\|f^{>d}\right\|_2^2. \tag{25}$$

To finish the proof, we show that $\left\|f^{>d}\right\|_2^2 \geqslant \Omega(\mathsf{var}(f))$. To do that, we upper-bound the weight of f on degrees 1 to d.

Let $g = f^{\leqslant d}$. We intend to bound $\mathsf{var}(g)$ using the Poincaré inequality, namely the first item in Fact A.2. Fix an order 1 Laplacian L_1. We have

$$\|\mathrm{L}_1 g\|_2^2 = \langle \mathrm{L}_1 g, \mathrm{L}_1 f\rangle \leqslant \|\mathrm{L}_1 g\|_4 \|\mathrm{L}_1 f\|_{4/3}. \tag{26}$$

As f is Boolean, $\mathrm{L}_1 f$ is $\{-1,0,1\}$-valued and so $\|\mathrm{L}_1 f\|_{4/3} = \|\mathrm{L}_1 f\|_2^{3/2}$, and next we bound $\|\mathrm{L}_1 g\|_4$. Note that

$$\|\mathrm{L}_1 g\|_4^4 = \mathop{\mathbb{E}}_{\substack{\mathrm{D}_1\\ \text{order 1 derivative}\\ \text{consistent with } \mathrm{L}_1}}\left[\|\mathrm{D}_1 g\|_4^4\right], \tag{27}$$

and we analyze $\|\mathrm{D}_1 g\|_4^4$ for all derivatives D_1. For that we use hypercontractivity, and we first have to show that $\mathrm{D}_1 g$ is global.

Fix a 1-derivative D_1, and set $h = \mathrm{D}_1 g$. By Lemma 8.10 (with $\tilde{f} = f - \mathbb{E}[f]$ instead of f), we get that for all $r\leqslant d-1$ and order r derivatives D we have

$$\begin{aligned}\|\mathrm{D}h\|_2 = \left\|\mathrm{D}\mathrm{D}_1\left(\tilde{f}^{\leqslant d}\right)\right\|_2 &\leqslant 2^{O(d^4)}\max_{\substack{\mathrm{D}' \text{ an } r-\text{derivative}\\ \mathrm{D}_1' \text{ a } 1-\text{derivative}}}\left\|\left(\mathrm{D}'\mathrm{D}_1'\tilde{f}\right)^{\leqslant d-r-1}\right\|_2 + n^{-10d}\left\|\tilde{f}^{\leqslant 4d}\right\|_2\\ &\leqslant 2^{-C\cdot d^4/2} + n^{-10d}\sqrt{\mathsf{var}(f)} \stackrel{def}{=} \delta,\end{aligned}$$

where we used $\mathrm{D}'\mathrm{D}_1'\tilde{f} = \mathrm{D}'\mathrm{D}_1' f$, which by assumption has 2-norm at most $2^{-C\cdot d^4}$, and $\left\|\tilde{f}^{\leqslant 4d}\right\|_2 \leqslant \left\|\tilde{f}\right\|_2 = \sqrt{\mathsf{var}(f)}$. For $r\geqslant d$, we have by Claim 4.3 that $\|\mathrm{D}h\|_2 = 0$. Thus, all derivatives of h have small 2-norm, and by Claim 4.2 we get that h is $(2d, 2^d\delta)$-global. Thus, from Theorem 1.4 we have that

$$\|\mathrm{D}_1 g\|_4 \leqslant 2^{O(d^3)}\delta^{1/2}\|\mathrm{D}_1 g\|_2^{1/2} \tag{28}$$

Plugging inequality (28) into (27) yields that

$$\|\mathrm{L}_1 g\|_4^4 \leqslant 2^{O(d^3)}\delta^2 \mathop{\mathbb{E}}_{\substack{\mathrm{D}_1\\ \text{order 1 derivative}\\ \text{consistent with } \mathrm{L}_1}}\left[\|\mathrm{D}_1 g\|_2^2\right] = 2^{O(d^3)}\delta^2\|\mathrm{L}_1 g\|_2^2 \leqslant 2^{O(d^3)}\delta^2\|\mathrm{L}_1 f\|_2^2.$$

Plugging this, and the bound we have on the $4/3$-norm $\mathrm{L}_1 f$, into (26), we get that

$$\|\mathrm{L}_1 g\|_2^2 \leqslant 2^{O(d^3)}\delta^{1/2}\|\mathrm{L}_1 f\|_2^2.$$

Summing this inequality over all 1-Laplacians and using Fact A.2, we get that

$$\mathsf{var}(g) \leqslant \frac{1}{n}\sum_{\mathrm{L}_1}\|\mathrm{L}_1 g\|_2^2 \leqslant 2^{O(d^3)}\delta^{1/2}\frac{2}{n-1}\sum_{\mathrm{L}_1}\|\mathrm{L}_1 f\|_2^2 = 2^{C\cdot d^3}\delta^{1/2} I[f]$$

for some absolute constant C, and we consider two cases.

The case that $I[f] \leqslant 2^{-C\cdot d^3}\delta^{-1/2}\mathsf{var}(f)/2$. In this case we get that $\mathsf{var}(g) \leqslant \mathsf{var}(f)/2$, and so $\left\|f^{>d}\right\| = \mathsf{var}(f) - \mathsf{var}(g) \geqslant \mathsf{var}(f)/2$. Plugging this into (25) finishes the proof.

The case that $I[f] \geqslant 2^{-C\cdot d^3}\delta^{-1/2}\mathsf{var}(f)/2$. By definition of δ we get that either $I[f] \geqslant 2^{C\cdot d^4/4}\mathsf{var}(f)$, in which case we are done, or $I[f] \geqslant 2^{-O(d^3)}n^{5d}\mathsf{var}(f)^{3/4}$, in which case we are done by the lower bound on n. $\square$