Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 133 (2023)

Product mixing in compact Lie groups

David Ellis* Guy Kindler† Noam Lifshitz‡ Dor Minzer§

Abstract

If G is a group, we say a subset S of G is *product-free* if the equation $xy = z$ has no solutions with $x, y, z \in S$. For $D \in \mathbb{N}$, a group G is said to be D-*quasirandom* if the minimal dimension of a nontrivial complex irreducible representation of G is at least D. Gowers showed that in a D-quasirandom finite group G, the maximal size of a product-free set is at most $|G|/D^{1/3}$. This disproved a longstanding conjecture of Babai and Sós from 1985.

For the special unitary group, $G = \mathrm{SU}(n)$, Gowers observed that his argument yields an upper bound of $n^{-1/3}$ on the measure of a measurable product-free subset. In this paper, we improve Gowers' upper bound to $\exp(-cn^{1/3})$, where $c > 0$ is an absolute constant. In fact, we establish something stronger, namely, *product-mixing* for measurable subsets of $\mathrm{SU}(n)$ with measure at least $\exp(-cn^{1/3})$; for this product-mixing result, the $n^{1/3}$ in the exponent is sharp.

Our approach involves introducing novel hypercontractive inequalities, which imply that the non-Abelian Fourier spectrum of the indicator function of a small set concentrates on high-dimensional irreducible representations. Our hypercontractive inequalities are obtained via methods from representation theory, harmonic analysis, random matrix theory and differential geometry. We generalize our hypercontractive inequalities from $\mathrm{SU}(n)$ to an arbitrary D-quasirandom compact connected Lie group for D at least an absolute constant, thereby extending our results on product-free sets to such groups.

We also demonstrate various other applications of our inequalities to geometry (viz., non-Abelian Brunn-Minkowski type inequalities), mixing times, and the theory of growth in compact Lie groups. A subsequent work due to Arunachalam, Girish and Lifshitz uses our inequalities to establish new separation results between classical and quantum communication complexity.

1 Introduction

A subset $\mathcal{A}$ of a group G is said to be *product-free* if $gh \notin \mathcal{A}$ for all $g, h \in \mathcal{A}$. The study of product-free subsets of groups has attracted significant attention over the past three

*Department of Mathematics, University of Bristol.

†Department of Computer Science, Hebrew University of Jerusalem.

‡Einstein Institute of Mathematics, Hebrew University of Jerusalem. Supported by the Israel Science Foundation (grant no. 1980/22).

§Department of Mathematics, Massachusetts Institute of Technology. Supported by a Sloan Research Fellowship, NSF CCF award 2227876 and NSF CAREER award 2239160.

ISSN 1433-8092

decades. In 1985, Babai and Sós [3] considered the problem of determining the largest size of a product-free set in a finite group G. They conjectured that exists an absolute positive constant $c_0 > 0$ such that any finite group G has a product-free set of size at least $c_0|G|$. In the Abelian case, this is quite easy to see, and had previously been observed by Erdős, in an unpublished communication to Babai and Sós. (In the cyclic case $(\mathbb{Z}_n, +)$, one can take a 'middle-third' construction, viz., $\{x \in \mathbb{Z}_n : \ n/3 < x \leqslant 2n/3\}$, as a large product-free set, and one can reduce to the cyclic case by observing that any finite Abelian group has a nontrivial cyclic quotient, and that the preimage of a product-free set under a quotient map is also product-free and of the same measure.) The exact answer in the Abelian case was given by Green and Ruzsa [15] in 2003: the largest product-free subset of a finite Abelian group G has size $c|G|$, where the function $c = c(G) \in [2/7, 1/2]$ was explicitly determined by Green and Ruzsa. The general Babai-Sós conjecture was disproved in 2008 by Gowers [14], who showed that if G is a finite group such that the minimal dimension of a nontrivial irreducible complex representation of G is equal to D, then any product-free subset of G has size at most $D^{-1/3}|G|$. It remains to observe that the quantity $D = D(G)$ is unbounded over finite non-Abelian groups G. For example, for the projective special linear group $\mathrm{PSL}_2(\mathbb{F}_q)$ (for q an odd prime power), we have $D(\mathrm{PSL}_2(\mathbb{F}_q)) = (q-1)/2$, so the measure of a product-free subset of $\mathrm{PSL}_2(\mathbb{F}_q)$ is at most $O(q^{-1/3})$, which tends to zero as q tends to infinity.

Gowers observed that his argument also implies that if G is an (infinite) compact group for which the minimal dimension of a nontrivial irreducible complex continuous representation is equal to D, then the maximal Haar measure of a measurable, product-free set in G is at most $D^{-1/3}$. For $\mathrm{SU}(n)$ we have $D(\mathrm{SU}(n)) = n$, implying an upper bound of $n^{-1/3}$ on the measure of a measurable product-free subset of $\mathrm{SU}(n)$. However, Gowers conjectured that for $\mathrm{SU}(n)$, the true answer is exponentially small in n. Indeed, as Gowers states, it seems difficult to come up with an example better than the following. Recall that group $\mathrm{SU}(n)$ acts on the complex unit sphere $\{v \in \mathbb{C}^n : \|v\|_2 = 1\}$, and take $\mathcal{A}$ to be the set of all matrices $A \in \mathrm{SU}(n)$ such that the real part of $\langle Ae_1, e_1\rangle$ is less than $-1/2$. As noted by Gowers, it follows from the triangle inequality that this set is product-free, and it is easy to check that the measure of $\mathcal{A}$ is $2^{-\Omega(n)}$.

In this work, we make progress towards proving Gowers' conjecture. Specifically, we improve Gowers' upper bound by a stretched exponential factor, viz., from $n^{-1/3}$ to $e^{-cn^{1/3}}$.

Theorem 1.1. *There exists an absolute constant $c > 0$ such that the following holds. Let $n \in \mathbb{N}$ and let $\mathcal{A} \subset \mathrm{SU}(n)$ be Haar-measurable and product-free. Then $\mu(\mathcal{A}) \leqslant \exp(-cn^{1/3})$.*

1.1 Quasirandomness for groups, and mixing.

Gowers' bound for product-free sets relies on a relationship between spectral gaps and dimensions of irreducible representations, a relationship which was first discovered by Sarnak and Xue [39]. In fact, Gowers' proof uses a beautiful connection between the problem and a purely representation-theoretic notion that Gowers called *quasirandomness* (due to a rough equivalence with the graph-quasirandomness of certain Cayley graphs, an equivalence which we shall explain below). For a group G we denote by $D(G)$ the minimal dimension of a non-trivial complex irreducible continuous representation of G. (Henceforth, for brevity, we will use the term *representation* to mean continuous representation.) For $d \in \mathbb{N}$, we say that a group G is *d-quasirandom* if $D(G) \geqslant d$.[1] Denoting by $\alpha(G)$ the largest possible density $\frac{|A|}{|G|}$ of a product-free set $A \subseteq G$ (if G is a finite group), Gowers showed that for any finite group G, $\alpha(G) \leqslant D(G)^{-1/3}$. Since $D(G)$ can be arbitrarily large (as is the case for the alternating groups, which have $D(A_n) = n - 1$ for all $n \geqslant 7$, and the groups $\mathrm{PSL}_2(\mathbb{F}_q)$ as mentioned above, and for many other natural infinite families of finite groups), this disproved the conjecture of Babai and Sós.

For finite groups, the quasirandomness parameter gives an almost complete description of the maximal size of a product free set. Pyber (see [14]) used the Classification of Finite Simple Groups to obtain a Kedlaya-type construction, showing that $\alpha(G) \geqslant D(G)^{-C}$ where $C > 0$ is an absolute constant. Nikolov and Pyber [33] later improved this to $\alpha(G) \geqslant \frac{1}{CD(G)}$. This established a remarkable fact, namely that the purely representation theoretic quasirandomness parameter $D(G)$ is polynomially related to the the combinatorial quantity $\alpha(G)$.

$$(CD(G))^{-1} \leqslant \alpha(G) \leqslant D(G)^{-1/3} \tag{1}$$

For compact connected Lie groups we obtain the following general variant of Theorem 1.1, which upper-bounds the size of a product-free sets in the group.

Theorem 1.2. *There exists an absolute constant $c > 0$ such that the following holds. Let G be a compact connected Lie group, and let $\tilde{G}$ be its universal cover. Let $\mathcal{A} \subset G$ be Haar-measurable and product-free. Let μ denote the Haar probability measure on G. Then $\mu(\mathcal{A}) \leqslant \exp(-cD(\tilde{G})^{1/3})$.*

An elegant argument of Gowers [14] (proof of Theorem 4.6, therein) for finite groups, which generalises very easily to the case of compact groups, shows that if G is a compact group then it has a measurable product-free subset of measure at least $\exp(-\Omega(D(G))$. In Section 2 we show that $D(G) = O(D(\tilde{G})^2)$. These two facts combine with Theorem 1.2 to give the following analogue of (1) for compact connected Lie groups.

Corollary 1.3. *There exists an absolute constant $c > 0$ such that the following holds. For every compact connected Lie group G,*

$$cD(G)^{1/6} \leqslant \log(1/\alpha(G)) \leqslant \frac{1}{c} D(G).$$

(We remark that our logs will always be taken with respect to the natural basis.) Corollary 1.3 says that, as with finite groups, the maximal measure of a measurable product-free set in a compact connected Lie group is controlled by the quasirandomness parameter, but this time the control moves to the exponent.

[1] To avoid confusion with the quasirandomness parameter for graphs, it might have been less ambiguous to call this notion 'd-group-quasirandomness', but as the latter is rather cumbersome we have opted for the above shorter formulation; we hope that this will not cause the reader confusion, in the sequel.

Quasirandomness (for groups) was a crucial ingredient in the 'Bourgain–Gamburd expansion machine', which is a three-step method for obtaining spectral gaps for Cayley graphs (see e.g. Tao [40], for an exposition). Briefly, this 'machine' proceeds as follows: one first shows that the graph has high girth, then one shows that there are no 'approximate subgroups' in which a random walk could be entrapped, and then quasirandomness is used (together with with the trace method) to finally obtain a spectral gap. Quasirandomness (for groups) has many other applications, such as in bounding the diameters of Cayley graphs (see e.g. the survery of Helfgott [18]).

The term 'quasirandomness' was used (for groups) by Gowers, due to the following connection with the (now classical) notion of quasirandomness for graphs. (There are, of course, now notions of quasirandomness for a huge variety of combinatorial and algebraic structures; roughly speaking, these say the structure behaves in a random-like way, in an appropriate sense.) We now need some more terminology. The *normalized adjacency matrix* $A_H \in \mathbb{R}^{V \times V}$ of a d-regular graph $H = (V, E)$ has (i, j)-th entry equal to $1/d$ if $\{i, j\} \in E$, and equal to zero otherwise. The graph H is said to be ε-*quasirandom* if all the nontrivial eigenvalues of A_H are at most ε in absolute value (here, 'nontrivial' means having an eigenvector orthogonal to the constant functions).

One of the striking consequences of d-quasirandomness for a finite group G, is that it implies that Cayley graphs of the form $Cay(G, S)$ are $(1/\text{poly}(d))$-quasirandom, whenever S is a dense subset of G. The fact that this only relies on density considerations and does not require any assumption on the structure of S, makes the notion of quasirandomness for groups rather powerful.

More generally, applications of quasirandomness for a group G can often be (re)phrased as follows. Suppose that G is d-quasirandom, and that we have a linear operator $T : L^2(G) \to L^2(G)$ whose nontrivial eigenvalues we want to bound (in absolute value) from above; suppose further that T commutes with either the left or the right action of G on $L^2(G)$. (In Gowers' proof, slightly rephrased, the operator T could be viewed as B^*B, where B is the bipartite adjacency matrix of the bipartite Cayley graph with vertex-classes consisting of two disjoint copies of G, and where the edges are all pairs of the form (g, sg) for $g \in G$ and $s \in S$, S being a product-free set in G.) Then by the commuting property, each eigenspace of T is a nontrivial representation of G, and therefore has dimension at least d; it follows that each nontrivial eigenvalue of G has multiplicity at least d. But the sum of the squares of the eigenvalues of T is equal to $\text{Trace}(T^2)$, and this yields the bound $d|\lambda|^2 \leqslant \text{Trace}(T^2)$ for all nontrivial eigenvalues λ of T. This is often called the Sarnak-Xue trick, as it was first employed in [39]

Bourgain and Gamburd used their 'expansion machine' (alluded to above) to show that taking two uniformly random elements $a, b \in \text{SL}_2(\mathbb{F}_p)$ is sufficient for the Cayley graph $\text{Cay}\left(\text{SL}_2(\mathbb{F}_p, \{a, b, a^{-1}, b^{-1}\}\right)$ to be an expander with high probability, p tending to infinity. It is a major open problem in the theory of Cayley graphs to obtain a similar result in the unbounded-rank case, for example for $\text{SL}_n(\mathbb{F}_p)$ where p is fixed and n tends to infinity. One of the properties that breaks down when one attempts to use the Bourgain–Gamburd expansion machine in the case of unbounded rank, is the dependence of the quasirandomness parameter on the cardinality of the group. Specifically, in order for the Bourgain–Gamburd expansion machine to work effectively for a group G, the quasirandomness parameter $D(G)$ needs to be polynomial in the cardinality of G. In the unbounded rank case, this no longer holds. For example, $D(SL_n(\mathbb{F}_p)) \leqslant p^n$ (consider the representation of dimension p^n induced by the natural action of $SL_n(\mathbb{F}_p)$ on $\mathbb{F}_p^n$). The situation is even worse for the alternating group A_n, as $D(A_n) = n-1$ for $n \geqslant 7$, and $n-1$ is less than logarithmic in the cardinality of the group.

1.2 Ideas and techniques

To improve on the upper bound of Gowers, we need to find methods for 'dealing with' the low-dimensional irreducible representations (more precisely, for dealing with the corresponding parts of the Fourier transform). In this paper, we develop some new techniques for this in the case of compact connected Lie groups. These techniques turn out also to be useful for finite groups; for example, in [27], analogues of some of our methods are developed for the alternating group A_n (where the idea of mixing is replaced by a refined notion, referred to therein as a 'mixing property for global sets').

Below we give indications of the new techniques that are used to obtain our improved bounds, and the various areas of mathematics from which they originate.

Level d inequalities and hypercontractivity

One of our key ideas is motivated by the (now well-developed) theory of the analysis of Boolean functions. A function

$$f\colon \{-1,1\}^n \to \mathbb{R}$$

has a *Fourier expansion* $f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S$, where $\chi_S : \{-1,1\}^n \to \{-1,1\}$ is defined by $\chi_S(x) := \prod_{i\in S} x_i$ for each $x \in \{-1,1\}^n$ and $S \subseteq [n]$. The functions χ_S, known as the *Fourier-Walsh functions* or *characters*, are orthonormal (with respect to the natural inner product on $\mathbb{R}[\{-1,1\}^n]$ induced by the uniform measure). The Fourier expansion gives rise to a coarser orthogonal decomposition, $f = \sum_{d=0}^{n} f^{=d}$, where

$$f^{=d} := \sum_{|S|=d} \hat{f}(S)\chi_S.$$

This is known as the *degree decomposition* (as each function $f^{=d}$ is a homogeneous polynomial of total degree d in the x_i's).

The *level d inequality* for the Boolean cube (essentially due to Kahn–Kalai–Linial [26] and Benjamini–Kalai–Schramm [6]) states that there exists an absolute constant $C > 0$, such that for a set $A \subseteq \{-1,1\}^n$ of density $\frac{|A|}{2^n} = \alpha$, if $d \leqslant \log(1/\alpha)$ then the characteristic function $f = 1_A$ satisfies $\|f^{=d}\|_2^2 \leqslant \alpha^2 \left(\frac{C\log(1/\alpha)}{d}\right)^d$. Roughly speaking, the level d inequality says that indicators of small sets are very much uncorrelated with low degree polynomials. One of our key ideas in this paper is to generalize the level d inequality from the Boolean cube to the setting of compact connected Lie groups.

The main tool in the proof of the Boolean level d inequality is the Bonami–Gross–Beckner hypercontactivity theorem. It states that the *noise operator* $\mathrm{T}_\rho f := \sum_{d=0}^{n} \rho^d f^{=d}$ is a contraction as an operator from L_q to L_p, for all $q > p \geqslant 1$ provided $0 \leqslant \rho \leqslant \sqrt{\frac{q-1}{p-1}}$. This immediately implies that $\|f^{=d}\|_q \leqslant \rho^{-d}\|f^{=d}\|_p$ for any function f. Roughly speaking, this last inequality says that L_p-norm of a low-degree function does not change too drastically with p. This is in stark contrast with the behaviour of indicator functions of small sets, $f = 1_A$. These satisfy $\|f\|_p = \alpha^{1/p}$, which does change rapidly with p. This difference in behaviours can be used to prove the level d-inequality, stating that indicators of small sets are essentially orthogonal to the low degree functions.

The same proof-concept works hand in hand with the representation theory of compact simple Lie groups. For simplicity, let us restrict our attention (at first) to the group $G = \mathrm{SO}(n)$. For each $d \in \mathbb{N} \cup \{0\}$, we let $V_{\leqslant d} \subseteq L^2(G)$ denote the subspace of $L^2(G)$ spanned by the polynomials of degree at most d in the matrix entries of $X \in G = \mathrm{SO}(n)$; so, for example, $X_{11}X_{22} \in V_{\leqslant 2}$. We also let $V_{=d} := V_{\leqslant d} \cap (V_{\leqslant d-1})^{\perp}$, for each $d \in \mathbb{N}$. Given $f \in L^2(G)$, we let $f^{\leqslant d}$ denote the orthogonal projection of f onto $V_{\leqslant d}$, and we let $f^{=d}$ denote the orthogonal projection of f onto $V_{=d}$, so that $f^{=d} = f^{\leqslant d} - f^{\leqslant d-1}$. The subspaces $V_{\leqslant d}$ and $V_{=d}$ are two-sided ideals of $L^2(G)$ (i.e., they are closed under both left and right actions of G on $L^2(G)$). Now, if J is a two-sided ideal of $L^2(G)$ and $T : L^2(G) \to L^2(G)$ is a linear operator that commutes with either the left or the right action of G (as will be the case with all the operators we will work with), it follows from the classical representation theory of compact groups (viz., the Peter-Weyl theorem and Schur's lemma) that T has J as an invariant subspace. Hence, such an operator T has each $V_{=d}$ as an invariant subspace, so each eigenspace of T can be taken to be within one of the $V_{=d}$'s. It therefore makes sense to consider quasirandomness relative to the degree decomposition. For each $d \in \mathbb{N}$, we let D_d be the smallest dimension of a subrepresentation of the G-representation $V_{=d}$. The obvious adaptation of the Sarnak-Xue trick, described above, then yields that for any eigenvalue λ of T with eigenspace within $V_{=d}$, we have $D_d|\lambda|^2 \leqslant \mathrm{Trace}(T^2)$. It turns out that D_d grows very fast with d, yielding very strong upper bounds on the corresponding $|\lambda|$ for large d.

On the other hand, an ideal level d inequality would imply that if A is an indicator of a small set, then most of its mass lies on the high degrees. This combines with the fast growth of D_d (with d) to give a much more powerful form of quasirandomness, one that takes into account the fact that f is $\{0, 1\}$-valued, and gives much better bounds.

We remark that the above degree decomposition can be easily extended to all compact linear Lie groups $G \leqslant GL_n(\mathbb{C})$ by letting $V_{\leqslant d}$ be the space of degree $\leqslant d$ polynomials in the real and imaginary parts of the matrix entries of $X \in G$. (In fact, this notion generalizes fairly easily to arbitrary compact simple Lie groups, even when they are not linear.) As in the $\mathrm{SO}(n)$ case, we let $f^{\leqslant d}$ denote the orthogonal projection of f onto $V_{\leqslant d}$.

We obtain the following level d inequality.

Theorem 1.4. *There exists absolute constants $c, C > 0$ such that the following holds. Let G be a simple compact Lie group equipped with its Haar probability measure μ. Suppose* *that $D(G) \geqslant C$. Let $A \subseteq G$ be a measurable set with $\alpha := \mu(A) \geqslant e^{-cD(G)}$. Then for each $d \in \mathbb{N} \cup \{0\}$ with $d \leqslant \log(1/\alpha)$, we have $\|f^{\leqslant d}\|_2^2 \leqslant \alpha^2 \left(\frac{2\log(1/\alpha)}{d}\right)^{Cd}$.*

When G is simply connected and $d \leqslant c\sqrt{n}$ we are able to obtain an even stronger level d inequality, which is similar to the one on the Boolean cube without the extra C factor in the exponent. This leads to the following.

Theorem 1.5. *There exists absolute constants $C, c > 0$ such that the following holds. Let G be a compact connected Lie group, let $\tilde{G}$ denote its universal cover, and write $n = D(\tilde{G})$. Suppose that $n \geqslant C$. Let $A \subseteq G$ be a measurable set with $\alpha := \mu(A) \geqslant \exp(-cn^{1/2})$. Then for each $d \in \mathbb{N} \cup \{0\}$ with $d \leqslant \log(1/\alpha)$, we have $\|f^{\leqslant d}\|_2^2 \leqslant \alpha^2 \left(\frac{C\log(1/\alpha)}{d}\right)^{d}$.*

It is this second level d-inequality that is responsible for the $1/3$ in the exponent of Theorem 1.2. Unfortunately, one would not be able to improve our $1/3$ in the exponent to the (conjectural) right one, merely by strengthening this level d-inequality. Indeed, our second level d inequality can be easily seen to be sharp up to the value of the absolute constant C, by considering sets of the form $\{A \in \mathrm{SO}(n) : \langle Ae_1, e_1\rangle > 1 - t\}$ for appropriate values of t, when $G = \mathrm{SO}(n)$, for example.

Both of our level d inequalities are inspired by the same ideas from the Boolean setting, together with an extra representation theoretic ideas. Namely, in order to show a level d inequality, we upper-bound q-norms of low degree polynomials in terms of their 2-norms, and then use Hölder's inequality. In the Boolean cube, such upper bounds follow from two facts. The first is that the noise operator T_ρ is hypercontractive. The second is that all the eigenvalues of the restriction of T_ρ to $V_{\leqslant d}$ are bounded from below by ρ^d. Our approach is to construct operators on $L^2(G)$ that satisfy the same two properties.

Differential geometry and Markov diffusion processes

Our level d inequalities stem from two techniques for obtaining hypercontractivity. Our first level d inequality, Theorem 1.4, is obtained via the following method. First, we observe that we assume without loss of generality that our group G is simply connected. (This is because every compact simple Lie group is a quotient of its universal cover by a discrete subgroup of its centre.) We then make use of classical lower bounds on the Ricci curvature of our (simply connected) compact simple Lie group. The Bakry-Emery criterion [4] translates such lower bounds on the Ricci curvature into log-Sobolev inequalities for the Laplace-Beltrami operator L. We then apply an inequality of Gross [16] to deduce a hypercontractive inequality for the operator e^{-tL} from the log-Sobolev inequality. This inequality then allows us to prove our first level d inequality. The operator e^{-tL} is the one corresponding to Brownian motion on G. In order to deduce our level d-inequality we rely on a formula for the eigenvalues of the Laplacian in terms of a step vector corresponding to each eigenspace. This formula is well-known in the theory of Lie groups; it is given for example in Berti and Procesi [7].

Random walks on bipartite graphs

There are two mutually adjoint linear operators that correspond to a random walk on a d-regular bipartite graph $B \subseteq L \times R$. We denote those by $T\colon L^2(L) \to L^2(R)$ and $T^*\colon L^2(R) \to L^2(L)$ and they are given by taking expectations over a random neighbour; explicitly, $(Tf)(x) = \mathbb{E}_{y\sim x} f(y)$ for $f \in L^2(L)$ and $x \in R$, and $(T^*g)(y) = \mathbb{E}_{x\sim y} g(x)$ for $g \in L^2(R)$ and $y \in L$. It is easy to see that both operators are contractions with respect to any norm. It turns out that given such a bipartite graph and given a hypercontractive operator S on R one gets for free that the operator T^*ST is hypercontractive. Filmus et al [11] used this idea to obtain a 'non-Abelian' hypercontractive estimate for 'global' functions on the symmetric group, from an 'Abelian' hypercontractive result for 'global' functions on $(\mathbb{Z}_n)^n$. (Informally, a 'global' function is one where one cannot increase the expectation very much by restricting the values of a small number of coordinates.)

In this work, we extend this idea to the continuous domain, by replacing a bipartite graph by a coupling of two probability distributions. Specifically, we consider the probability space $(\mathbb{R}^{n\times n}, \gamma)$ of n by n Gaussian matrices (i.e., $\mathbb{R}^{n\times n}$ with each entry being an independent standard Gaussian), and the Haar measure on $\mathrm{O}(n)$. For $(\mathbb{R}^{n\times n}, \gamma)$, the Ornstein–Uhlenbeck operator U_ρ is a hypercontractive analogue of the noise operator from the Boolean case. We couple $(\mathbb{R}^{n\times n}, \gamma)$ with $\mathrm{SO}(n)$ by applying the Gram–Schmidt operation on the columns of a given Gaussian matrix (flipping the sign of the last column, if necessary, so as to ensure that the determinant is equal to one). We note that essentially the same coupling has been used before, e.g. by Jiang [22]; however, it has not been used before (to our knowledge) to analyse the distribution of high-degree polynomials in the matrix-entries, which is crucial in our work.

This coupling gives rise to operators $\mathrm{T}_{\mathsf{col}}$ and $\mathrm{T}^*_{\mathsf{col}}$, similar to the ones in the discrete case. The hypercontractive inequality for the Ornstein–Uhlenbeck operator U_ρ, together with our coupling implies a hypercontractive inequality for the operator $\mathrm{T}'_\rho := \mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}}$. We then use a symmetrization trick to obtain an operator $\mathrm{T}_\rho := \mathbb{E}_{B\sim\mu} R_B^* \mathrm{T}'_\rho R_B$, where R_B corresponds to right multiplication by B. The symmetrization does not change the hypercontractive properties, which are the same as for U_ρ (see Theorem 8.8), but it has the advantage of allowing us to analyse more easily the eigenvalues of the operator.

Representation theory

The hypercontractive inequality for the operator T_ρ is useful due to the fact that it immediately gives bounds on the norms of eigenfunctions of T_ρ. Because of the symmetrization, T_ρ commutes with the action of G from both sides. Therefore, the Peter-Weyl theorem implies that every isotypical[2] component of $L^2(G)$ is contained in an eigenspace of T_ρ.

We eventually show that the eigenvalues of the restriction of T_ρ to V_d are at least $(c\rho)^d$, for some absolute constant $c > 0$. This implies that T_ρ is indeed a good analogue of the noise operator on the Boolean cube, and of the Ornstein–Uhlenbeck operator U_ρ, i.e.

[2] If ρ is an irreducible representation of G and V is a G-module, the *ρ-isotypical component* of V is the sum of all subrepresentations of V that are isomorphic to ρ.

the noise operator on Gaussian space. We obtain this lower bound by showing that each isotypical component contains certain functions that are nice to deal with, functions we call the *comfortable juntas*.

The latter are defined as follows. We define a *d-junta* to be a function in the matrix entries of $X \in \mathrm{SO}(n)$ that depends only upon the upper-left d by d minor of X. Such a d-junta is said to be *comfortable* d-junta if it is contained in the linear span of the monomials $\{m_\sigma : \sigma \in S_d\}$, where $m_\sigma : \mathrm{SO}(n) \to \mathbb{R}$ is defined by $m_\sigma(X) = \prod_{i=1}^d X_{i,\sigma(i)}$ for each $X \in \mathrm{SO}(n)$, for each permutation $\sigma \in S_d$.

Random matrix theory

One of the main discoveries of random matrix theory is that the entries of a random orthogonal matrix behave (in an appropriate sense) like independent Gaussians of the same expectation and variance: at least, when one restricts minors of the matrix that are not too large. (In fact, this holds for many different models of random matrices, not just the orthogonal ensemble.) The power of this discovery is of course that a Gaussian random matrix is *a priori* much easier to analyse than e.g. the random matrix given by the Haar measure on a group.

One way to test that two distributions are similar is to apply a continuous 'test function' and take expectations. Usually, for applications in random matrix theory, the test function can be taken to be an arbitrary fixed polynomial.

When computing the eigenvalues of our operator T_ρ we need to show a similarity in distribution between the upper $d \times d$-minor of $\mathrm{O}(n)$ and the $d \times d$ minor of a random Gaussian matix. For us, however, it is not sufficient to look at a single polynomial of fixed degree. Instead, we need to show a similarity in the distribution with respect to our comfortable d-juntas (where d may be as large as $\sqrt{n}$, rather than an absolute constant). Hence, while the philosophy is similar to that of random matrix theory, we require new techniques enabling us to deal with the distributions of polynomials whose degrees may be a function of n, indeed up to $\sqrt{n}$.

1.3 Applications

In this section we list several applications of our hypercontractive theory: to some problems in group theory, in geometry, and in probability.

To state some of our results, we need some more terminology. If G is a compact connected Lie group, we define $n(G) := D(\tilde{G})$, where $\tilde{G}$ denotes the universal cover of G. It is well-known that, for each $m \in \mathbb{N}$, we have $D(\mathrm{SU}(m)) = D(\mathrm{Spin}(m)) = m$ and $D(\mathrm{Sp}(m)) = 2m$ (and all these groups are simply connected except for $\mathrm{Spin}(2)$); we also have $D(\mathrm{SO}(m)) = m$. Since $\mathrm{Spin}(m)$ is the universal cover of $\mathrm{SO}(m)$ for all $m > 2$, we have $n(\mathrm{SO}(m)) = m$ for all $m > 2$. As we will see in the next section, any compact connected semisimple Lie group G with $D(G)$ at least an absolute constant, can be written in the form $(\prod_{i=1}^r K_i)/F$ where each K_i is one of $\mathrm{SU}(n_i), \mathrm{Spin}(n_i)$ or $\mathrm{Sp}(n_i)$ for some $n_i \geqslant 3$, and F is a finite subgroup of the centre of $\prod_{i=1}^r K_i$; the universal cover of such is $\prod_{i=1}^r K_i$, and $D(\prod_{i=1}^r K_i) = \Theta(\min_i n_i)$. Hence, the quantity $n(G)$ has a very explicit description in terms of the structure of the Lie group G.

Growth in groups: the diameter problem

The theory of growth in groups has been a very active area of study in recent decades, and an important class of problem in this area is to determine the diameter of a metric space defined by a group (e.g., the diameter of a Cayley graph of the group). For a compact group G equipped with its Haar probability measure, and a measurable generating set $\mathcal{A} \subseteq G$ of measure μ, it is natural to consider the metric space on G where the distance between x and y is defined to be the minimal length of a word in the elements of $\mathcal{A}$ and their inverses which is equal to xy^{-1}. The diameters of such metric spaces in the case where G is finite have become a focus of intense study in the last two decades: see e.g. the works of Liebeck and Shalev [30], Helfgott [17], Helfgott and Seress [19], Pyber and Szabo [35] and Breuillard, Green and Tao [8].

For a subset $\mathcal{A}$ of a group G and $t \in \mathbb{N}$, we define

$$\mathcal{A}^t := \{ a_1 \cdot a_2 \cdots a_t \mid a_1, a_2, \dots, a_t \in \mathcal{A}\} .$$

The *diameter problem for G with respect to $\mathcal{A}$* asks for the smallest positive integer t for which $\mathcal{A}^t = G$. For a compact group G and a real number $0 < \alpha \leqslant 1$, the *diameter problem for sets of measure α in G* asks for the minimum possible diameter of a measurable set in G of measure α.

In the case where G is a compact and connected group, we note that the diameter of G with respect to any subset $\mathcal{A}$ of positive measure is finite. This follows (almost) immediately from Kemperman's theorem [28], which states that for any compact connected group G (equipped with its Haar probability measure μ) and any measurable $\mathcal{A}, \mathcal{B} \subset G$, we have $\mu(\mathcal{A}\mathcal{B}) \geqslant \min\{\mu(\mathcal{A}) + \mu(\mathcal{B}), 1\}$.

We make the following conjecture, concerning the diameter of large sets.

Conjecture 1.6. *Let G be one of* $\mathrm{SU}(n), \mathrm{SO}(n), \mathrm{Spin}(n)$ *or* $\mathrm{Sp}(n)$*, and let $\mathcal{A} \subseteq G$ be a measurable subset of measure ν. Then the diameter of G with respect to $\mathcal{A}$ is $O(\nu^{-1/(\ell n)})$, where $\ell = 1$ in the case of* $\mathrm{SO}(n)$ *and* $\mathrm{Spin}(n)$*, $\ell = 2$ in the case of* $\mathrm{SU}(n)$*, and $\ell = 4$ in the case of* $\mathrm{Sp}(n)$*. In particular, if $\nu \geqslant e^{-cn}$, then the diameter of G with respect to $\mathcal{A}$ is at most $O_c(1)$.*

We note that if true, the conjecture is essentially tight, as can be seen for $\mathrm{SO}(n)$ by considering the set

$$\mathcal{S}_\varepsilon := \{X \in \mathrm{SO}(n) \colon \text{the angle between } Xe_1 \text{ and } e_1 \text{ is at most } \varepsilon\},$$

For $\varepsilon \leqslant 1/2$, we have $\mu(\mathcal{S}_\varepsilon) = (\Theta(\varepsilon))^n$, and the diameter of $\mathrm{SO}(n)$ with respect to $\mathcal{S}_\varepsilon$ is $\Theta(1/\varepsilon)$. If $\pi : \mathrm{Spin}(n) \to \mathrm{SO}(n)$ is the usual (double) covering homomorphism, then the lift $\pi^{-1}(\mathcal{S}_\varepsilon)$ is a subset of $\mathrm{Spin}(n)$ of the same measure as $\mathcal{S}_\varepsilon$ (using, of course, the Haar probability measure on both groups), and the diameter of $\mathrm{Spin}(n)$ with respect to $\pi^{-1}(\mathcal{S}_\varepsilon)$

is the same the diameter of $\mathrm{SO}(n)$ with respect to $\mathcal{S}_\varepsilon$, since $\pi(\mathcal{A}^t) = (\pi(\mathcal{A}))^t$ for any subset $\mathcal{A} \subset \mathrm{Spin}(n)$ and any $t \in \mathbb{N}$. Hence, $\pi^{-1}(\mathcal{S}_\varepsilon)$ demonstrates tightness for $\mathrm{Spin}(n)$. The group $\mathrm{SU}(n)$ acts transitively on the unit sphere in $\mathbb{C}^n$, which can be identified with S^{2n-1}, and the group $\mathrm{Sp}(n)$ acts transitively on the unit sphere in $\mathbb{H}^n$, which can be identified with S^{4n-1}; both actions are angle-preserving (in S^{2n-1} and S^{4n-1} respectively). So our above construction for $\mathrm{SO}(n)$ (which comes from the action of $\mathrm{SO}(n)$ on S^{n-1}) has the obvious analogues for $\mathrm{SU}(n)$ and $\mathrm{Sp}(n)$, which we conjecture are sharp for those groups.

We show that for a compact connected Lie group G with $n(G) = n$, for all $\delta > 0$ and all measurable subsets $\mathcal{A}$ of G with measure at least $2^{-cn^{1-\delta}}$, the diameter of G with respect to $\mathcal{A}$ is at most $O_\delta(1)$.

Theorem 1.7. *For each $\delta > 0$ there exist $n_0, k > 0$ such that the following holds. Let $n > n_0$ and let G a compact connected Lie group with $n(G) = n$. If $\mathcal{A} \subset G$ is a Haar-measurable set, and $\mu(\mathcal{A}) \geqslant 2^{-n^{1-\delta}}$, then $\mathcal{A}^k = G$.*

Doubling inequalities for groups

Theorem 1.7 follows from a new lower bound on $\mu(\mathcal{A}^2)$, where $\mathcal{A} \subseteq G$ is a measurable subset of the compact connected Lie group G. We prove the following 'doubling inequality'.

Theorem 1.8. *There exists absolute constants $C, c > 0$ such that the following holds. Let G be a compact connected Lie group with $n(G) = n \geqslant C$. Let $\mathcal{A} \subseteq G$ be a measurable set with $\mu(\mathcal{A}) \geqslant e^{-cn}$. Then $\mu(\mathcal{A}^2) \geqslant \mu(\mathcal{A})^{0.1}$.*

The problem of giving a lower bound on $\mu(\mathcal{A}^2)$ in terms of $\mu(\mathcal{A})$, for $\mathcal{A}$ a measurable subset of a compact group G, dates back to the work of Henstock and Macbeath [20] from 1953, the aforemenentioned bound of Kemperman [28] from 1964, and the work of Jenkins [21] from 1973. Several recent works of Jing, Tran and Zhang have introduced some powerful new methods into the field. For instance, in [24], Jing, Tran and Zhang generalized the Brunn-Minkowskii inequality from $\mathbb{R}^n$ to an arbitrary connected Lie group, using the Iwasawa decomposition to facilitate an inductive approach; their result is essentially sharp for helix-free Lie groups. In [25], they used techniques from O-minimal geometry to show that that $\mu(\mathcal{A}^2) \geqslant 3.99\mu(\mathcal{A})$ for all measurable subsets $\mathcal{A} \subseteq \mathrm{SO}(3)$ of sufficiently small measure. In a forthcoming paper [23] they prove that there exists a function $\delta = \delta(n)$ and an absolute constant $c > 0$, such that if $\mathcal{A} \subseteq \mathrm{SO}(n)$ is a measurable set of measure at most $\delta(n)$, then $\mu(\mathcal{A}^2) \geqslant 2^{cn^{1/10}}\mu(\mathcal{A})$; the function $\delta(n)$ satisfies $\delta(n) \leqslant 2^{-n^{1+c'}}$ where $c' > 0$ is an absolute constant. (For comparison, we note that Theorem 1.8, in conjunction with Theorem 1.10 below, imply the existence of an absolute constant $c > 0$ such that $\mu(\mathcal{A}^2) \geqslant \min\{2^{cn^{1/2}}\mu(\mathcal{A}), 0.99\}$ for all measurable subsets $\mathcal{A} \subset \mathrm{SO}(n)$ of measure at least 2^{-cn}, so our result and that of Jing, Tran and Zhang leave a 'gap' between them.) It remains an open problem to determine whether $\mu(\mathcal{A}^2) \geqslant \min\{2^{n/10}\mu(\mathcal{A}), 0.99\}$ for all subsets $\mathcal{A} \subset \mathrm{SO}(n)$.

Spectral gaps

We also give the following upper bound on the spectral gaps of the operator corresponding to convolution by $\frac{1_A}{\mu(A)}$. If G is a compact group equipped with its (unique) Haar probability measure μ, for a measurable set $\mathcal{A} \subset G$ we write $x \sim \mathcal{A}$ to mean that x is chosen (conditionally) according to the Haar measure μ, conditional on the event that $x \in \mathcal{A}$.

Theorem 1.9. *There exist absolute constants $c, C > 0$ such that the following holds. Let G be compact connected Lie group and suppose $n := n(G) \geqslant C$. Let $\mathcal{A} = \mathcal{A}^{-1}$ be a symmetric, measurable set in G and suppose that $\mu(\mathcal{A}) \geqslant e^{-cn^{1/2}}$. Then the nontrivial spectrum of the operator T defined by $Tf(x) = \mathbb{E}_{a\sim A}[f(ax)]$ is contained in the interval*

$$\left[-\sqrt{\frac{C\log 1/\alpha}{n}}, \sqrt{\frac{C\log 1/\alpha}{n}}\right].$$

Mixing times

Let G be a compact group, equipped with its (unique) Haar probability measure; then every measurable subset $\mathcal{A} \subseteq G$ of positive Haar measure corresponds to a random walk on G. Indeed, we may define a (discrete-time) random walk $R_{\mathcal{A}} = (X_t)_{t\in\mathbb{N}\cup\{0\}}$ on G, by letting $X_0 = \mathrm{Id}$, and for each $t \in \mathbb{N}$, if $X_{t-1} = x$ then $X_t = ax$, where a is chosen uniformly at random from $\mathcal{A}$. In the case where G is finite and $\mathcal{A}$ is closed under taking inverses, this is the usual random walk associated to the Cayley graph $\mathrm{Cay}(G, \mathcal{A})$. One of the fundamental problems associated to such random walks is to determine their *mixing time*. (Following Larsen and Shalev [29], we say that the *mixing time* of a Markov chain $(X_t)_{t\in\mathbb{N}\cup\{0\}}$ is the minimal non-negative integer T such that the total variation distance between the distribution of X_T and the uniform distribution, is at most $1/e$. We note that $1/e$ could be replaced by any other absolute constant $c \in (0, 1)$, without materially altering the definition; Larsen and Shalev use the constant $1/e$ as it makes the statement of certain results concerning S_n and A_n more elegant.)

Larsen and Shalev [29] considered the case where $\mathcal{A}$ is a normal set, i.e. a set closed under conjugation, and G is the alternating group A_n. They showed that for each $\varepsilon > 0$, if $\mathcal{A} \subseteq A_n$ of density $\frac{2|\mathcal{A}|}{n!} \geqslant \exp\left(-n^{1/2-\varepsilon}\right)$, then the mixing time of $R_{\mathcal{A}}$ is 2, provided that $n \geqslant n_0(\varepsilon)$ is sufficiently large depending on ε. Their proof was based upon a heavy use of character theory. Their result is almost sharp, in the sense the number $1/2$ cannot be replaced by any number smaller than $1/2$. We show that a similar phenomenon holds for compact connected Lie groups, even when $\mathcal{A}$ is not a normal set.

Theorem 1.10. *There exist absolute constants $c, n_0 > 0$, such that the following holds. Let G be a compact connected Lie group with $n := n(G) > n_0$. Let $\mathcal{A} \subseteq G$ be a measurable set with Haar measure at least $e^{-cn^{1/2}}$. Then the mixing time of the random walk $R_{\mathcal{A}}$ is 2.*

This result is essentially best possible. For instance, taking $G = \mathrm{SO}(n)$, we may take $\mathcal{A} = \{X \in \mathrm{SO}(n) : X_{11} > 10/n^{1/4}\}$. It is easy to see that the mixing time of $R_{\mathcal{A}}$ is 3, while $\mu(\mathcal{A}) = \exp(-\Theta(n^{1/2}))$.

Product mixing

Gowers' proof of his upper bound on the sizes of product-free sets actually establishes a stronger phenomenon, known as *product mixing*. We say that a compact group G (equipped with its Haar probability measure μ) is an (α, ε)-*mixer* if for all sets $A, B, C \subseteq G$ of Haar probability measures $\geqslant \alpha$, when choosing independent uniformly random elements $a \sim A$ and $b \sim B$, the probability that $ab \in C$ lies in the interval $(\mu(C)(1-\varepsilon), \mu(C)(1+\varepsilon))$. Gowers' proof actually yields the following statement: there exists an absolute constant $C > 0$, such that if G is a D-quasirandom compact group, then it is a $(CD^{-1/3}, 0.01)$-mixer. (The proof is given only for finite groups, but it generalises easily to all compact groups.) For finite groups, Gowers' product-mixing result is sharp up to the value of the constant C. Here, we obtain an analogous result for compact connected Lie groups, where the $n^{-1/3}$ moves to the exponent.

Theorem 1.11. *For any $\varepsilon > 0$, there exist $c, n_0 > 0$ such that the following holds. Let $n > n_0$ and let G be a compact connected Lie group with $n := n(G) > n_0$. Set $\alpha = \exp(-cn^{-1/3})$. Then G is an (α, ε)-mixer.*

This result is sharp up to the dependence of the constants $c = c(\varepsilon)$ and $n_0 = n_0(\varepsilon)$ upon ε. Indeed, we may take $G = \mathrm{SO}(n)$ and let $\mathcal{A} = \mathcal{B} = \{X \in \mathrm{SO}(n) : \ X_{11} > 10/n^{1/3}\}$ and $\mathcal{C} = \{X \in \mathrm{SO}(n) : \ X_{11} < -10/n^{1/3}\}$, to obtain a triple of sets each of measure $e^{-\Theta(n^{1/3})}$, such that when choosing $a \sim A$ and $b \sim B$ independently, the probability that $ab \in C$ is smaller than $\frac{1}{2}\mu(C)$.

Homogeneous dynamics and equidistribution

Suppose that a compact Lie group G acts on a topological space X. The space X is said to be G-*homogeneous* if G acts transitively and continuously on X (the latter meaning that the action map from $G \times X$ to X is continuous); in this case, X has a unique G-invariant probability measure, which is called the Haar probability measure. We obtain the following equidistribution result for homogeneous spaces.

Theorem 1.12. *For each $\varepsilon > 0$ there exist $c, n_0 > 0$ such that the following holds. Let G be a compact connected Lie group with $n(G) =: n > n_0$. Let X be a G-homogeneous topological space, and let μ_X denotes its unique G-invariant (Haar) probability measure. Suppose that $\mathcal{A} \subseteq G$ and $\mathcal{B} \subseteq X$ are both measurable sets of Haar probability measures $\geqslant e^{-cn^{1/2}}$. Let ν be the probability measure on X which is given by the distribution of $a(b)$, for a uniform random $a \sim \mathcal{A}$ and an (independent) uniform random $b \sim \mathcal{B}$. Then the total variation distance between μ and ν is less than ε.*

L^q-norms of low degree polynomials

We obtain the following upper bounds on the q-norms of low degree polynomials (we state the result for $\mathrm{SO}(n)$, for simplicity).

Theorem 1.13. *There exist absolute constants $c, C > 0$ such that the following holds. Let $q > 2$ and let $f \in L^2(\mathrm{SO}(n))$ be a polynomial of degree d in the matrix entries of $X \in \mathrm{SO}(n)$. If $d \leqslant cn$, then*

$$\|f\|_{L^q(\mu)} \leqslant q^{Cd} \|f\|_{L^2(\mu)}.$$

If moreover, $d \leqslant c\sqrt{n}$, then

$$\|f\|_{L^q(\mu)} \leqslant (C\sqrt{q})^d \|f\|_{L^2(\mu)}.$$

Separation of quantum and classical communication complexity

Starting with their introduction to computer science in the seminal paper of Kahn Kalai and Linial [26], hypercontractive inequalities have found a huge number of applications in various branches of computer science and related fields (see e.g. [13, 31, 10, 36], to name but a few). While these applications have generally required hypercontractive inequalities for functions on discrete sets, some applications require continuous domains. For example, in the paper of Klartag and Regev [37], a hypercontractive inequality for functions on the n-sphere is used to obtain a lower bound on the number of (classical) communication bits required for two parties to jointly compute a certain function. While with quantum communication, the value of that function can be computed by one party transmitting only $O(\log n)$ quantum-bits to the other, it was shown in [37] that classical communication requires at least $\Omega(n^{1/3})$ bits of communication to be sent, even if parties are allowed to send bits both ways, thereby showing an exponential separation between the power of classical communication and that of one-way quantum communication.

In the field of quantum communication, establishing a significant separation between classical communication and practically implemented modes of quantum communication remains a major open problem. In a forthcoming work, Arunachalam, Girish, and Lifshitz [2] apply our hypercontractive inequality for $\mathrm{SU}(n)$ to make a substantial step towards this goal. They used it to obtain an exponential separation between classical communication and a more realistic version of quantum communication, namely the one-clean-qubit model.

2 Preliminaries: Quasirandomness and min-rank

In this section we show that, for D at least an absolute constant, the universal cover $\tilde{G}$ of a D-quasirandom compact connected Lie group G is a product of 'classical' (compact, simple, simply connected) Lie groups of the form $\mathrm{Spin}(n), \mathrm{SU}(n), \mathrm{Sp}(n)$. We then make use of this to determine $D(\tilde{G})$, and we show that $D(\tilde{G}) \leqslant 4D(G)^2$.

In what follows, as usual, a *compact group* G is a Hausdorff topological group for which the group operations (or equivalently, the map $(g,h) \mapsto gh^{-1}$) are continuous. We recall that a compact group has a unique left-multiplication-invariant probability measure (called the Haar measure), which is also the unique right-multiplication-invariant probability measure. As usual, if G is a compact group, we let

$$L^2(G) = \{f : G \to \mathbb{C} : \ \mathbb{E}_\mu[|f|^2] < \infty\}/\sim,$$

where the expectation is with respect to the Haar probability measure μ on G and the equivalence relation $\sim$ is defined by $f \sim g$ iff $f = g$ μ-almost-everywhere, and we view $L^2(G)$ as a Hilbert space, with the natural inner product,

$$\langle f, g \rangle := \mathbb{E}_\mu[\overline{f}g].$$

We make use of the following fact, appearing for example in [34] Chapter 10, Section 7.2, Theorem 4, page 380. (Note that the word 'Lie' is missing from the statement of this theorem; this omission is clearly just a typographical error.)

Fact 2.1. *Every compact connected Lie group is Lie-isomorphic to a group of the form* $(\prod_{i=1}^r K_i \times T)/F$*, where each* K_i *is a simply connected simple compact Lie group (equivalently,* K_i *is one of* $\mathrm{Sp}(n_i)$ *for some* $n_i \geqslant 1$*,* $\mathrm{Spin}(n_i)$ *for some* $n_i \geqslant 3$*,* $\mathrm{SU}(n_i)$ *for some* $n_i \geqslant 2$*, or the compact form of one of the five exceptional Lie groups, for each* $i \in [r]$*),* T *is a finite-dimensional torus (i.e.* $T = (\mathbb{R}/\mathbb{Z})^m$ *for some integer* m*), and* F *is a finite group contained in the center of* $\prod_{i=1}^r K_i \times T$*, with* $F \cap T = \{1\}$*.*

Lemma 2.2. *Suppose that* $D > 1$ *and that* G *is a* D*-quasirandom compact connected Lie group. Then* G *is semisimple.*

Proof. Write $G = (\prod_{i=1}^r K_i \times T)/F$, where each K_i is a simply connected simple compact Lie group, T is a finite-dimensional torus (i.e. $T = (\mathbb{R}/\mathbb{Z})^m$ for some integer m), and F is a finite group contained in the center of $\prod_{i=1}^r K_i \times T$, with $F \cap T = \{1\}$, as in the above fact. Semisimplicity of G is equivalent to $T = \{1\}$. Suppose on the contrary that $T \neq \{1\}$. Let π be the projection map from $\prod_{i=1}^r K_i \times T$ onto the T component. Since $F \cap T = \{1\}$, we have $\pi(F) = \{1\}$, so the projection π induces a (surjective) group homomorphism $\tilde{\pi}$ from G to T, and therefore G has a quotient isomorphic to $(\mathbb{R}/\mathbb{Z})^m$ for some integer $m \geqslant 1$; any nontrivial complex one-dimension irreducible representation of the latter quotient lifts to one of G, contradicting the D-quasirandomness of G (for any $D > 1$) and proving the lemma. □

We also recall the following standard fact.

Fact 2.3. *Every compact semisimple Lie group has finite centre.*

We now show that if G is sufficiently quasirandom, then the exceptional groups do not make an appearance as some K_i when writing $G = (\prod_{i=1}^r K_i)/F$.

Lemma 2.4. *Set* $D_0 = 248$*. Let* G *be a compact connected Lie group, and suppose that it is* D*-quasirandom for some* $D > D_0$*. Then* $G = (\prod_{i=1}^r K_i)/F$*, where each* K_i *is one of* $\mathrm{Sp}(n_i), \mathrm{Spin}(n_i), \mathrm{SU}(n_i)$ *for some* $n_i \geqslant \sqrt{D}/2$ *and* F *is a subgroup of the (finite) centre of* $\prod_{i=1}^r K_i$*.*

Proof. By the previous lemma, provided $D_0 > 1$, G is semisimple. Hence, we may write $G = (\prod_{i=1}^r K_i)/F$, where each K_i is one of $\mathrm{Sp}(n_i)$, $\mathrm{Spin}(n_i)$, $\mathrm{SU}(n_i)$ or the compact form of one of the five exceptional Lie groups, for each i, and F is a finite group contained in the (finite) center of $\prod_{i=1}^r K_i$. As the quotient of a D-quasirandom group is D-quasirandom, we may project to any one of the components and still obtain a D-quasirandom group K_i/F'. (In detail, let π_i denote projection of $\prod_{j=1}^r K_j$ onto the K_i factor; π_i induces a surjective homomorphism from G onto $K_i/\pi_i(F)$, and since F is a subgroup of the centre of $\prod_{j=1}^r K_j$, F_i is a subgroup of the centre of K_i. The group $K_i/\pi_i(F)$ is therefore a quotient of G, and so inherits its D-quasirandomness.) It is therefore sufficient to consider the case where $G = K_1/F'$. We now note that the adjoint representation of K_1 factors through K_1/F' (since F' is contained in the centre of K_1), so it can also be viewed as a representation of K_1/F'. As the adjoint representation of K_1 is not a sum of copies of the trivial representation (this follows from the fact that K_1 is non-Abelian), its dimension (which is the same as the dimension of the Lie group K_1) is at least D. The five exceptional Lie groups, E_6, E_7, E_8, F_4 and G_2, have dimensions 78, 133, 248, 52 and 14 respectively, so K_1 cannot equal any of these (since $D > D_0 = 248$). Hence, K_1 is one of $\mathrm{Sp}(n_1)$, $\mathrm{Spin}(n_1)$ or $\mathrm{SU}(n_1)$. The dimensions of these Lie groups are $n_1(2n_1+1)$, $n_1(n_1-1)/2$ and n_1^2-1 respectively, so we obtain $n_1(2n_1+1) \geqslant D$, which implies that $n_1 \geqslant \sqrt{D}/2$. This completes the proof of the lemma. □

The following (non-standard) definition will be convenient for us.

Definition 2.5. *Let G be a compact, connected, semisimple Lie group and write $G = (\prod_{i=1}^r K_i)/F$, where, as above, K_i is one of* $\mathrm{Sp}(n_i)$, $\mathrm{Spin}(n_i)$ *or* $\mathrm{SU}(n_i)$ *for each i, and F is a finite subgroup of the (finite) centre of G. We define the* min-rank *of G to be* $\min\{n_1, \ldots, n_r\}$.

Using this terminology, the above lemma can be restated by saying that if a compact connected Lie group G is D-quasirandom for large enough D, then it has min-rank at least $\sqrt{D}/2$.

We remark that the *rank* of a Lie group is defined to be the dimension of any one of its Cartan subgroups, so the ranks of $\mathrm{Sp}(n_i)$, $\mathrm{Spin}(n_i)$ and $\mathrm{SU}(n_i)$ are respectively n_i, $\lfloor n_i/2 \rfloor$ and $n_i - 1$, so in particular are all $\Theta(n_i)$; hence, while the min-rank of G is not exactly the minimum of the ranks of the K_i's (where the K_i's are as above), it is within an absolute constant factor thereof. (We hope this slight abuse of terminology will not cause confusion.)

To establish that $D(\tilde{G}) \leqslant 4D(G)^2$ we also need the following.

Lemma 2.6. *Let G be a simply compact semisimple Lie group of min-rank m. Then its universal cover $\tilde{G}$ satisfies $D(\tilde{G}) \in \{m, 2m\}$.*

Proof. Write $G = \prod_{i=1}^r K_i/F$. As the projection map from $\prod_{i=1}^r K_i$ to G is a cover map, and since $\prod_{i=1}^r K_i$ is simply connected, we obtain that $\tilde{G} = \prod_{i=1}^r K_i$. The lemma now follows from the fact that the complex irreducible representations of a product $\prod_{i=1}^r K_i$ of finitely many compact groups are tensor products of complex irreducible representations, of the form $\rho_1 \otimes \ldots \otimes \rho_r$ where ρ_i is an complex irreducible representation of ρ_i, for each i together with the fact that $D(\mathrm{SU}(n)) = D(\mathrm{SO}(n)) = n, D(\mathrm{Sp}(n)) = 2n$. □

Lemma 2.6 shows that, when proving the theorems in the introduction, we may replace $D(\tilde{G})$ with the min-rank of G.

3 Good groups and fine groups

In this section we define some basic properties of compact connected groups, which we later use to prove various growth properties. We define *graded* and *strongly quasirandom* groups, and *hypercontractive* groups; we say that groups satisfying all these properties are *good.* We also define a somewhat weaker (or, technically, incomparable) notion of a *fine* group.

The compact, simple, simply connected real Lie groups of large enough rank, i.e SU(n), Sp(n) and Spin(n), are indeed good (this is proved in Section 8). We show that goodness is preserved when taking products and quotients (quotients, that is, by closed normal subgroups, as usual), thereby showing that every D-quasirandom group is a good graded group, provided D is sufficiently large.

Definition 3.1 (Graded groups)**.** *For $n \in \mathbb{N}$, we say a compact connected group G is* n-graded *if there exists an orthogonal direct sum,*

$$L^2(G) = \bigoplus_{d=0}^{\lceil n/2\rceil-1} V_{=d} \oplus V_{\geqslant n/2},$$

such that the spaces $V_{=d}$ are invariant under the action of G from both sides, and where $V_{=0}$ contains only the constant functions. For an n-graded group and an integer $0 \leqslant d_0 < n/2$, we denote by $V_{>d_0}$ the direct sum $V_{>d_0} := \bigoplus_{d=d_0+1}^{\lceil n/2\rceil-1} V_{=d} \oplus V_{\geqslant n/2}$. (Note that we will sometimes write $V^G_{=d}$ in place of $V_{=d}$, when we want to stress that the group in question is G, e.g. if there are several groups involved in our argument.)

Remark 3.2. *Note that it follows from the definition of an n-grading that $V_{\geqslant n/2}$ is also invariant under the action of G from both sides.*

Grading for the compact simply connected simple Lie groups (of large enough rank). We note in this section that SO(n), SU(n), Sp(n) and Spin(n) are all n-graded, for $n \geqslant 3$. For the group SO(n), for each integer $0 \leqslant d < n/2$ we define $V_{\leqslant d}$ to be the subspace of $L^2(\mathrm{SO}(n))$ spanned by degree $\leqslant d$ multivariate polynomials in the matrix entries of $X \in \mathrm{SO}(n)$ (so, for example, $V_{\leqslant 2}$ contains the polynomial $X_{11}X_{12}$). For notational convenience, for $0 < r < d/2$ we define $V_{<r}$ to be $V_{\leqslant d}$, where d is the maximal integer less than r. We then set $V_{=0} := V_{\leqslant 0}$ and $V_{=d} := V_{\leqslant d} \cap (V_{\leqslant d-1})^{\perp}$ for each integer $1 \leqslant d < n/2$, and we define $V_{\geqslant n/2} := (V_{<n/2})^{\perp}$. Note that each $V_{=d}$ is finite-dimensional, but $V_{\geqslant n/2}$ is infinite-dimensional.

For the special unitary group we perform a similar construction, except that one views the complex entries of the input matrix as a pair of real numbers. More precisely, we define $V_{\leqslant d}$ to consist of the functions f that are degree $\leqslant d$ multivariate polynomials in the real and imaginary parts of the matrix entries of $X \in \mathrm{SU}(n)$ (so, for example, $V_{\leqslant 3}$ contains the polynomial $\mathrm{Re}(X_{11})\mathrm{Im}(X_{11})\mathrm{Re}(X_{12})$). As in the SO($n$) case, we then set $V_{=0} := V_{\leqslant 0}$ and $V_{=d} := V_{\leqslant d} \cap (V_{\leqslant d-1})^{\perp}$ for each integer $1 \leqslant d < n/2$, and we set $V_{\geqslant n/2} := (V_{<n/2})^{\perp}$.

For the compact symplectic group, $\mathrm{Sp}(n)$, we view it as the group of n by n unitary matrices over the field of quaternions (see Section 6.5 for more details), and we define $V_{\leqslant d}$ to be the vector subspace of $L^2(\mathrm{Sp}(n))$ spanned by degree $\leqslant d$ multivariate polynomials in the real parts, the **i**-parts, the **j**-parts and the **k**-parts of the matrix entries; we then proceed as in the previous two cases.

For the spin group $\mathrm{Spin}(n)$ things are a little more involved, as it has no straightforward description as a linear group. In order to define the grading we make use of the double covering homomorphism $\pi\colon \mathrm{Spin}(n) \to \mathrm{SO}(n)$ (recall that $\mathrm{Spin}(n)$ is the universal covering group of $\mathrm{SO}(n)$, and that this cover is a double cover, for each $n \geqslant 3$). The covering homomorphism π gives rise to an embedding $i\colon L^2(\mathrm{SO}(n)) \to L^2(\mathrm{Spin}(n))$ given by $if = f \circ \pi$. We then take the grading of the spin group to be

$$\mathrm{Spin}(n) = \bigoplus_{d<n/2} i(V_{=d}^{\mathrm{SO}(n)}) \oplus (i(V_{<n/2}^{\mathrm{SO}(n)}))^{\perp},$$

i.e. $V_{=d}^{\mathrm{Spin}(n)} = i(V_{=d}^{\mathrm{SO}(n)})$ for each $0 \leqslant d < n/2$, and $V_{\geqslant n/2}^{\mathrm{Spin}(n)} = (i(V_{<n/2}^{\mathrm{SO}(n)}))^{\perp}$.

We remark that the above arguments also imply that $\mathrm{SO}(n)$, $\mathrm{SU}(n)$, $\mathrm{Sp}(n)$ and $\mathrm{Spin}(n)$ are m-graded for all $m \in \mathbb{N}$ and all $n \geqslant 3$, but we will only require the n-grading, in the sequel.

Next we define a notion of 'strong quasirandomness' for graded compact groups. In section 6 we show that the simply connected n-graded groups are all c-strongly-quasirandom for some absolute constant $c > 0$.

Definition 3.3 (Strongly quasirandom graded group). *We say an n-graded compact group G is $\left((Q_d)_{d=0}^{\lceil n/2\rceil-1}, Q\right)$-*strongly-quasirandom *if the minimal dimension of a subrepresentation of $V_{=d}$ (as a left G-module) is $\geqslant Q_d$ for all integers $0 \leqslant d \leqslant \lceil n/2\rceil - 1$, and is $\geqslant Q$ for $V_{\geqslant n/2}$.*

For $c > 0$ we say that the n-graded compact group G is c-strongly-quasirandom *if it is $((Q_d)_{d=1}^{\lceil n/2\rceil-1}, Q)$-strongly-quasirandom when we set $Q_d := \left(\frac{cn}{d}\right)^d$ for $d < cn/(1+c)$, and $Q, Q_d := (1+c)^{cn/(1+c)}$ for $d \geqslant cn/(1+c)$.*

In Section 3.4, we show that all the (infinite families of) compact simply connected simple Lie groups are c-strongly quasirandom for some absolute constant $c > 0$.

Theorem 3.4. *The n-graded compact groups $\mathrm{SU}(n), \mathrm{Sp}(n), \mathrm{Spin}(n)$ (for $n \geqslant 3$) are all c-strongly quasirandom for some absolute constant $c > 0$, when equipped with our chosen n-grading.*

Definition 3.5 (Beckner operator for graded groups). *Let G be an n-graded compact group, let r be an integer with $0 \leqslant r < n/2$, and let $0 \leqslant \delta \leqslant 1$. We define the* Beckner operator *$T_{\delta,r} : L^2(G) \to L^2(G)$ by $T_{\delta,r}(f) := \sum_{i=0}^{r} \delta^i f^{=i}$, for all $f \in L^2(G)$.*

Definition 3.6 (Hypercontractive group). *Let $C > 0$ and let r be an integer with $0 \leqslant r < n/2$. We say that an n-graded group compact G is (r, C)-*hypercontractive *if for every $q \geqslant 2$ and every $0 \leqslant \delta \leqslant 1/(C\sqrt{q})$, we have $\|T_{\delta,r}\|_{2\to q} \leqslant 1$.*

The following is an easy consequence of hypercontractivity.

Lemma 3.7. *Let G be an n-graded (r, C)-hypercontractive group, where $0 \leqslant r < n/2$. Then for every integer $d \leqslant r$, every $q \geqslant 2$ and function $f \in V_{=d}$, we have*

$$\|f\|_q \leqslant \left(C^2 q\right)^{d/2} \|f\|_2.$$

Proof. Let $f \in V_{=d}$. Then $\delta^d \|f\|_q = \|T_{\delta,r} f\|_q \leqslant \|f\|_2$ for all $\delta \leqslant 1/(C\sqrt{q})$; setting $\delta = 1/(C\sqrt{q})$ completes the proof. □

In Section 8 we show that the compact simple simply connected n-graded Lie groups $\mathrm{Sp}(n), \mathrm{SU}(n)$ and $\mathrm{Spin}(n)$ are $(c\sqrt{n}, C)$-hypercontractive for some positive absolute constants C and c, when they are eqipped with our chosen n-grading.

Theorem 3.8. *The groups $\mathrm{Sp}(n), \mathrm{SU}(n), \mathrm{Spin}(n)$ are $(c\sqrt{n}, C)$-hypercontractive, when equipped with our chosen n-grading, for some positive absolute constants C and c.*

Definition 3.9 (Good groups)**.** *An n-graded compact group G is said to be* (C, c)-good *if it is $(cn^{1/2}, C)$-hypercontractive and c-strongly quasirandom.*

The next theorem follows from Theorem 3.8 and Theorem 3.4.

Theorem 3.10. *For each $n \geqslant 3$, the n-graded groups $\mathrm{Sp}(n), \mathrm{SU}(n), \mathrm{Spin}(n)$ (equipped with our choice of n-grading) are (C, c)-good, for some absolute positive constants C and c.*

We prove Theorem 3.10 in Section 8.

3.1 Fine groups

We also have a closely related notion of a fine group. It is a weaker form of hypercontractivity, but we manage to prove it for higher values of d, viz., up to linear in n.

Definition 3.11 (Weakly hypercontractive group)**.** *Let $C > 1$ and $1 \leqslant r < n/2$. An n-graded compact group G is* (r, C)-weakly hypercontractive *if for every function $f \in L^2(G)$ and every $q \geqslant 2$ and $0 \leqslant \delta \leqslant 1/q^C$ we have*

$$\|T_{\delta,r} f\|_q \leqslant \|f\|_2.$$

The following lemma follows similarly to Lemma 3.7.

Lemma 3.12. *Let G be an n-graded (r, C)-weakly hypercontractive compact group, where $C > 1$ and $1 \leqslant r < n/2$. Then for any integer $d \leqslant r$, every $q \geqslant 2$, and any function $f \in V_{=d}$, we have*

$$\|f\|_q \leqslant q^{Cd} \cdot \|f\|_2$$

Definition 3.13 (Fine groups)**.** *If $c > 0$ and $C > 1$, an n-graded compact group G is said to be* (C, c)-fine *if it is both (cn, C)-weakly hypercontractive, and c-strongly-quasirandom.*

In Section 7 we show that the (infinite families of) compact simply connected simple groups, viz. $\mathrm{Sp}(n), \mathrm{SU}(n)$, and $\mathrm{Spin}(n)$, are all fine, as n-graded groups equipped with our chosen gradings.

Theorem 3.14. *For $n \geqslant 3$, the n-graded groups $\mathrm{Sp}(n), \mathrm{SU}(n)$, and $\mathrm{Spin}(n)$ are (C, c)-fine, for some absolute constants $C > 1$ and $c > 0$, when equipped with our chosen n-gradings.*

3.2 Goodness and fineness are preserved under taking products and quotients

In this subsection, we show that goodness and fineness are preserved under taking products, and quotients (quotients, that is, by closed normal subgroups) — more precisely, that suitable gradings can be defined on products and quotients. This, together with Theorems 3.10 and 3.14, implies that all compact Lie groups of large-enough min-rank are both good and fine.

Lemma 3.15. *Assuming Theorem 3.10, the following holds. There exist positive constants c, C and n_0 such that if $n > n_0$, then every compact connected Lie group of min-rank n can be equipped with an n-grading that makes it (C, c)-good, as an n-graded group.*

Lemma 3.16. *Assuming Theorem 3.14, the following holds. There exist positive constants c, C and n_0 such that if $n > n_0$, then every compact connected Lie group of min-rank n can be equipped with an n-grading that makes it (C, c)-fine, as an n-graded group.*

Grading for group products. Let G and H be compact groups, and let f and g be functions on G and H respectively. We write $f \otimes g$ for the function on $G \times H$ given by $(x, y) \mapsto f(x)g(y)$. If U is a closed linear subspace of $L^2(G)$ and V is a closed linear subspace of $L^2(H)$, then we denote by $U \otimes V$ be the Hilbert-space tensor product of U and V, i.e. the linear subspace of $L^2(G \times H)$ consisting of the closure of the linear span of the set of functions $\{f \otimes g : f \in U, g \in V\}$. We recall that if $\{u_i\}_{i=1}^{\infty}$ is a Hilbert-space basis for U and $\{v_i\}_{i=1}^{\infty}$ is a Hilbert-space basis for V, then $\{u_i \otimes v_j\}_{i,j=1}^{\infty}$ is a Hilbert-space basis for $U \otimes V$.

Definition 3.17. *Let $n \geqslant m$, let G be an n-graded compact group and H be an m-graded compact group. We give $G \times H$ the structure of an m-graded compact group by setting*

$$V_{=d}^{G \times H} = \bigoplus_{d_1 + d_2 = d} V_{=d_1}^{G} \otimes V_{=d_2}^{H}$$

for $d < m/2$, and $V_{\geqslant m/2}^{G \times H} = (V_{\leqslant \lceil m/2 \rceil - 1}^{G \times H})^{\perp}$.

For products of more than two compact groups, $G_1 \times G_2 \times \ldots \times G_\ell$ say, we simply iterate the above definition (noting associativity). Viz., if G_i is n_i-graded for $i = 1, 2, \ldots, \ell$, then we let $n = \min\{n_1, \ldots, n_\ell\}$ and we give $G_1 \times \ldots \times G_\ell$ the structure of an n-graded compact group by setting

$$V_{=d}^{G_1 \times \ldots \times G_\ell} = \bigoplus_{d_1 + \ldots + d_\ell = d} V_{=d_1}^{G_1} \otimes \ldots \otimes V_{=d_\ell}^{G_\ell}$$

for $d < n/2$, and $V_{\geqslant n/2}^{G_1 \times \ldots \times G_\ell} = (V_{\leqslant \lceil n/2 \rceil - 1}^{G_1 \times \ldots \times G_\ell})^\perp$.

Our goal is now to show that the product of good groups is good. We make use of the following lemma of Beckner [5].

Lemma 3.18. *Let $X_1, \ldots, X_r, Y_1, \ldots, Y_r$ be probability spaces. Let $T_1, \ldots T_r$ be linear operators such that $T_i \colon L^2(X_i) \to L^q(Y_i)$ for all $i \in [r]$. Then $\|T_1 \otimes \cdots \otimes T_r\|_{2 \to q} \leqslant \prod_{i=1}^r \|T_i\|_{2 \to q}$.*

Lemma 3.19. *Let $G_1, \ldots, G_l$ be compact groups and suppose that each G_i is n_i-graded. Write $n = \min\{n_1, \ldots, n_l\}$. Suppose that each G_i is (r, C)-hypercontractive, where $r \in \mathbb{N}$ with $r < n/2$; then $G := \prod_{i=1}^l G_i$ (with the above n-grading) is also (r, C)-hypercontractive.*

Proof. Let $q \geqslant 2$, and let $\delta = \frac{1}{C\sqrt{q}}$. Let $T_i \colon L^2(G_i) \to L^2(G_i)$, $T \colon L^2(G) \to L^2(G)$ be the appropriate $T_{\delta,r}$ Beckner operators. Then by hypothesis $\|T_i\|_{2 \to q} \leqslant 1$ for all $i \in [r]$; our goal is to show that $\|T\|_{2 \to q} \leqslant 1$. This now follows from Lemma 3.18 and the fact that T can be decomposed as $T = T_1 \otimes \cdots \otimes T_l \circ S$, where S is given by $f \mapsto f^{\leqslant r}$. We have

$$\|Tf\|_q \leqslant \|T_1 \otimes \cdots \otimes T_l\|_{2 \to q} \|Sf\|_2 \leqslant \|f\|_2$$

for each f. □

A similar argument works for weakly hypercontractive groups, yielding the following lemma.

Lemma 3.20. *Let $G_1, \ldots, G_l$ be compact groups and suppose that each G_i is n_i-graded. Write $n = \min\{n_1, \ldots, n_l\}$. Suppose that each G_i is (r, C)-weakly hypercontractive, where $0 \leqslant r < n/2$. Then the n-graded group $G := \prod_{i=1}^l G_i$ (with the above n-grading) is also (r, C)-weakly hypercontractive.*

Grading for quotients. We now define a grading for quotients of graded groups, and show that hypercontractivity (and weak hypercontractivity) is preserved under taking quotients, using this 'induced' grading.

Definition 3.21. *Let G be an n-graded compact group, let H be a closed normal subgroup of G, and let $\pi \colon G \to G/H$ be the quotient map. The n-graded structure on $L^2(G/H)$ induced from that on $L^2(G)$ is given by letting $V_{=d}^{G/H}$ consist of all functions f in $L^2(G/H)$ such that $f \circ \pi$ is in $V_{=d}^G$. The space $V_{\geqslant n/2}^{G/H}$ is defined similarly.*

We note that this definition is consistent with our choices of the gradings of $\mathrm{Spin}(n)$ and of $\mathrm{SO}(n)$.

Lemma 3.22. *The subspaces $V_{=d}^{G/H}$ constitute an n-grading of the compact group G/H.*

Proof. We first note that if

$$L^2(G) = \left(\bigoplus_{d=0}^{\lceil n/2 \rceil - 1} V_{=d}^G \right) \oplus V_{\geqslant n/2}$$

is a grading of $L^2(G)$, then each $V_{=d}^G$ is closed (being an orthogonal complement of a subspace), as is $V_{\geqslant n/2}^G$.

Let $i\colon L^2(G/H) \to L^2(G)$ be given by $i(f) = f \circ \pi$. We let i^* be its adjoint, which is given explicitly by $i^*(f)(xH) = \mathbb{E}_{h \in H}[f(xh)]$, where the expectation is taken with respect to the Haar probability measure on H. Now i^* commutes with the action of G (from either the left or the right) and therefore the spaces $i^*(V_{=d}^G)$ and $i^*(V_{\geqslant n/2}^G)$ are invariant under both the left and the right actions of G. Since i^* preserves orthogonality, these spaces are also pairwise orthogonal. Since $i^* \circ i$ is the identity (so $i^*(L^2(G)) = L^2(G/H)$), we obtain that the spaces $i^*(V_{=d}^G), i^*(V_{>n/2})$ constitute a grading of $L^2(G/H)$. The fact that $i^* \circ i$ is the identity also imiplies that $V_{=d}^{G/H} \subseteq i^*(V_{=d}^G)$ for each $d < n/2$, and that $V_{\geqslant n/2}^{G/H} \subseteq i^*(V_{\geqslant n/2}^G)$. We now claim that $i^*(V_{=d}^G) = V_{=d}^{G/H}$ for all $d < n/2$, and that $i^*(V_{\geqslant n/2}^G) = V_{\geqslant n/2}^{G/H}$. To prove this, it suffices to show that $V_{=d}^G$ is invariant under $i \circ i^*$ for each $d < n/2$, and that $V_{\geqslant n/2}^G$ is invariant under $i \circ i^*$. (Indeed, the latter implies that $i^*(V_{=d}^G) \subseteq V_{=d}^{G/H}$ for each $d < n/2$, and that $i^*(V_{\geqslant n/2}^G) \subseteq V_{\geqslant n/2}^{G/H}$.) Now, $i \circ i^*$ is given by $f \mapsto (x \mapsto \mathbb{E}_{h \sim H}[f(xh)])$. Suppose that $f \in V_{=d}^G$. By the right-invariance, each function $f(xh)$ is then in the space $V_{=d}^G$, and as $V_{=d}^G$ is closed, it follows that the average $i \circ i^*$ also lies in $V_{=d}^G$. Exactly the same argument works with $V_{\geqslant n/2}^G$, proving the claim. This completes the proof of the lemma. □

Lemma 3.23. *Let G be an n-graded compact group, and let H be a closed normal subgroup of G. Suppose that G is (r, C)-(weakly) hypercontractive, where $0 \leqslant r < n/2$. Then G/H is also (r, C)-(weakly) hypercontractive, when equipped with the induced n-grading defined above.*

Proof. Let $i\colon L^2(G/H) \to L^2(G)$ be given by $i(f) = f \circ \pi$, as before. We notice that by definition of the grading and of the Beckner operator, it holds that $T_{\delta,r}^G \circ i = i \circ T_{\delta,r}^{G/H}$. Noting that the map i is an L^p-isometric embedding for all p, we have that for any $f \in L^2(G/H)$,

$$\left\| T^{G/H} f \right\|_q = \left\| i \circ T^{G/H} f \right\|_q = \left\| T^G(i(f)) \right\|_q \leqslant \| i(f) \|_2 = \| f \|_2 .$$

This completes the proof of the lemma. □

If the map i from earlier induces an isomorphism between the spaces $V_{=d}^{G/H}$ and $V_{=d}^G$ for all $d \leqslant r$, then the converse of Lemma 3.23 also holds; this will be useful for going from $\mathrm{SO}(n)$ to $\mathrm{Spin}(n)$.

Lemma 3.24. *Let G be an n-graded compact group, and let H be a closed normal subgroup of G. Equip G/H with the induced (quotient) grading, defined above. Suppose that G/H is (r,C)-(weakly) hypercontractive as an n-graded group, where $0 \leqslant r < n/2$. Suppose further that for all $d \leqslant r$, the gradings satisfy $i\left(V_{=d}^{G/H}\right) = V_{=d}^{G}$. Then G is also (r,C)-(weakly) hypercontractive, as an n-graded group.*

Proof. Let us denote the Beckner operator $T_{\delta,r}$ on $L^2(G)$ by T, and the corresponding operator on $L^2(G/H)$ by T'. Then we may write $T \circ i = i \circ T'$. Composing with i^* we obtain

$$T \circ i \circ i^* = i \circ T' \circ i^*.$$

We now claim that $T \circ i \circ i^* = T$. First note that each space $V_{=d}^G, V_{>r}^G$ is $i \circ i^*$-invariant. We can therefore use fact that the operator T annihilates $V_{>r}^G$ to deduce that the operator $T \circ i \circ i^*$ agrees with T on $V_{>r}^G$. Our claim will follow once we show that $i \circ i^*$ is the identity on $V_{\leqslant r}^G$. To accomplish that we note that i is injective, and by the hypothesis the restriction of i to the corresponding $V_{=d}$ spaces is also surjective, and thus so is its restriction to $V_{\leqslant r}^G$. As the operator $i^* \circ i$ is the identity we obtain that the restriction of $i \circ i^*$ to $V_{\leqslant r}^G$ is the identity as well.

We can therefore write $T = i \circ T' \circ i^*$, and using the fact that i is an L^q-isometric embedding and that i^* contracts 2-norms we obtain:

$$\|T\|_{2\to q} = \|T' \circ i^*\|_{2\to q} \leqslant \|T'\|_{2\to q} \|i^*\|_{2\to 2} \leqslant 1.$$

□

Conclusion. We have shown that (weak) hypercontractivity is preserved under taking products and quotients. It is easy to check that c-strong-quasirandomness is also preserved under taking products or quotients (using the gradings above), so we immediately obtain Lemma 3.15 and Lemma 3.16.

4 Growth in good groups

In this section we prove Theorems 1.2, 1.5, 1.9, 1.10 1.11 and 1.12. For now, the reader may consider the objective of proving Theorem 1.2 as motivation for what follows.

Let G be a compact group, and let μ be the Haar probability measure on G. We would like to bound $\mu(A)$ for a set $A \subseteq G$ that is product free. We first note that the property of being product free can be stated in terms of convolutions.

Definition 4.1. *For two functions $f, g \in L^2(G)$, we define their* convolution $f * g \in L^2(G)$ *by*

$$f * g(x) := \int f(xy^{-1}) g(y) d\mu(y).$$

For $f \in L^2(G)$, we write T_f for the linear operator from $L^2(G)$ to itself defined by $g \mapsto g * f$. Observe that if $A \subset G$ is a product-free set of density $\mu(A) = \alpha$, and $f = \frac{1_A}{\alpha}$, then $\langle T_f 1_A, 1_A \rangle = 0$. If G is an n-graded group, we can decompose $g := 1_A$ into its projections to the $V_{=d}$'s, and write $g = \sum_{d=0}^{\lceil n/2 \rceil - 1} g^{=d} + g^{\geqslant n/2}$, where $g^{=d}$ is the projection of g onto $V_{=d}$. Noting that $g^{=0} \equiv \alpha$, this allows us to expand

$$\langle T_f 1_A, 1_A \rangle = \langle T_f g, g \rangle$$

as a sum of a main term, α^2, and other terms of the form $\langle T_f g^{=d}, g^{=d} \rangle$ or $\langle T_f g^{\geqslant n/2}, g^{\geqslant n/2} \rangle$. We upper-bound each term by using Cauchy-Schwarz:

$$|\langle T_f g^{=d}, g^{=d} \rangle| \leqslant \|T_f\|_{V_{=d}} \|g^{=d}\|_2^2,$$

where for a closed subspace $M \leqslant L^2(G)$ and a linear operator $T : L^2(G) \to L^2(G)$ we write $\|T\|_M$ for the supremum of $\frac{\|Tv\|_2}{\|v\|_2}$ over all nonzero $v \in M$.

Our goal will be to show that these other term make a negligible contribution to the sum compared to the main term α^2. We accomplish that by observing that the space $V_{=d}$ is T_f-invariant. This shows that the operator $T_f^* T_f$ can be diagonalized inside $V_{=d}$. It also implies that $\|T_f\|_{V_{=d}}^2$ is equal to the maximal eigenvalue of $T_f^* T_f$. We then upper bound the maximal eigenvalue of $T_f^* T_f$ inside $V_{=d}$, showing that these eigenvalues get smaller and smaller as the degree gets larger. Finally, we combine our upper bound on the eigenvalues of T_f with a level d-inequality which shows that the L^2-mass of g is concentrated on the high degrees. Together, we obtain that the sum of terms $\langle T_f g^{=d}, g^{=d} \rangle$ is indeed negligible.

Our upper bound on the 'degree d' eigenvalues of $T_f^* T_f$ follows by combining a level d inequality with a lower bound on the dimension of each eigenspace of $T_f^* T_f$. We use the fact that each such eigenspace is a sub-representation of $V_{=d}$ and therefore by strong quasirandomness must have dimension $\geqslant Q_d$, for the appropriate quasirandomness parameter Q_d. We upper bound $|\langle T_f g^{\geqslant n/2}, g^{\geqslant n/2} \rangle|$ in a similar fashion.

4.1 Level d inequalities and the eigenvalues of T_f

Recall that the *Hilbert-Schmidt norm* of a linear operator T on a separable Hilbert space H is defined by

$$\|T\|_{\mathrm{HS}}^2 := \sum_{i=1}^{\infty} \|T(e_i)\|_2^2,$$

where $\{e_i\}_{i=1}^{\infty}$ is any Hilbert-space basis for H; if T is a compact operator, then $\|T\|_{\mathrm{HS}}$ is the square root of the sum of the eigenvalues of T^*T (counted with multiplicity). One standard fact (see e.g. [9], page 267) is the following.

Fact 4.2. *Let $f \in L^2(G)$ and let T_f be the linear operator from $L^2(G)$ to itself defined by $g \mapsto g * f$. Then T_f is a compact operator, and the Hilbert–Schmidt norm of T_f is equal to the 2-norm of f:*

$$\|T_f\|_{HS} = \|f\|_2.$$

We recall our notation for the norm of an operator on a subspace of $L^2(G)$.

Definition 4.3. *Let $M \leqslant L^2(G)$ be a closed subspace. Let $T\colon L^2(G) \to L^2(G)$ be a linear operator; then we write $\|T\|_M := \sup\{\|Tf\|_2/\|f\|_2 :\ f \in M \setminus \{0\}\}$.*

We now give our upper bound on the level d eigenvalues of T_f.

Lemma 4.4. *Let G be an n-graded $((Q_d)_{d=1}^{n/2}, Q)$-quasirandom group. Let $f \in L^2(G)$. Then the spaces $V_{=d}, V_{\geqslant n/2}$ are T_f-invariant. Moreover, $\|T_f\|_{V_{=d}} \leqslant \frac{\|f^{=d}\|_2}{\sqrt{Q_d}}$ and $\|T_f\|_{V_{\geqslant n/2}} \leqslant \frac{\|f\|_2}{\sqrt{Q}}$.*

Proof. We first claim that the subspaces $V_{=d}, V_{\geqslant n/2}$ are all T_f invariant. To see this, observe that if $U \leqslant L^2(G)$ is a closed subspace that is invariant under the right-action of G, then for every $g \in U$, we have $g * f \in U$ as well. (Indeed, let $h \in U^{\perp}$; then

$$\langle g * f, h\rangle = \int\int g(xy^{-1})f(y)h(x)d\mu(y)d\mu(x) = \int\int g(xy^{-1})f(y)h(x)d\mu(x)d\mu(y) = 0,$$

using Fubini and the fact that for each fixed $y \in G$ the function $x \mapsto g(xy^{-1})f(y)$ lies in U. Hence, $g * f \in (U^{\perp})^{\perp} = U$.) Applying this with $U = V_{=d}$, which is a closed subspace invariant under the right action of G, we see that the spaces $V_{=d}$ are indeed T_f-invariant. Similarly, applying it with $U = V_{\geqslant n/2}$ (which is also a closed subspace invariant under the right action of G), we see that $V_{\geqslant n/2}$ is also T_f-invariant.

Fix $d < n/2$, and let us orthogonally decompose f as $f = f^{=d} + f'$. Then $T_f = T_{f^{=d}} + T_{f'}$, by the linearity of convolution. We now assert that T_f agrees with $T_{f^{=d}}$ on $V_{=d}$. Essentially the same argument as that above shows that if $U \leqslant L^2(G)$ is a closed subspace that is invariant under the left-action of G, then $g * f' \in U$ for every $f' \in U$ and $g \in L^2(G)$; applying this with $U = V_{=d}^{\perp}$, we obtain that $T_{f'}g \in V_{=d}^{\perp}$ for every $g \in L^2(G)$. Hence, if $g \in V_{=d}$, then $T_{f'}g = 0$, proving our assertion. It follows that $\|T_f\|_{V_{=d}}^2$ is the maximal eigenvalue of the operator $T_{f^{=d}}^* T_{f^{=d}}$. On the other hand, Fact 4.2 implies that $\|f^{=d}\|_2^2$ is the sum of the eigenvalues of $T_{f^{=d}}^* T_{f^{=d}}$ counted with multiplicity. To complete the proof that $\|T_f\|_{V_{=d}} \leqslant \frac{\|f^{=d}\|_2}{\sqrt{Q_d}}$ we show that each such multiplicity is $\geqslant Q_d$ and so

$$Q_d\|T_f\|_{V_{=d}}^2 \leqslant \|f^{=d}\|_2^2.$$

To lower-bound the multiplicities of the eigenvalues of $T_{f^{=d}}^* T_{f^{=d}}$ inside $V_{=d}$ we note that $T_{f^{=d}}^* T_{f^{=d}}$ commutes with the left-action of G, since $T_{f^{=d}}$ does. This implies that the eigenspaces of $T_{f^{=d}}^* T_{f^{=d}}$ are left G-submodules of $V_{=d}$. Each such submodule contains an irreducible representation, which has dimension $\geqslant Q_d$ by hypothesis. The proof that $\|T_f\|_{V_{\geqslant n/2}} \leqslant \frac{\|f\|_2}{\sqrt{Q}}$ is similar. □

We now upper-bound $\|T_f\|_{V_{=d}}$ by proving a corresponding level d-inequality.

Theorem 4.5. *Let G be an (r, C)-hypercontractive group. Let $f\colon G \to \{0,1\}$ and write $\alpha := \mathbb{E}[f]$. Let $d \in \mathbb{N}$ be such that $0 < d \leqslant \min\{\frac{1}{2}\log(1/\alpha), r\}$. Then*

$$\|f^{=d}\|_2^2 \leqslant \left(\frac{10C}{d}\right)^d \alpha^2 \log^d(1/\alpha).$$

Proof. Let $q = \frac{\log(1/\alpha)}{d}$. Let q' be the Hölder conjugate of q, i.e. the number satisfying $\frac{1}{q} + \frac{1}{q'} = 1$. Then by (r, C)-hypercontractivity and Lemma 3.7 we have

$$\|f^{=d}\|_2^2 = \langle f, f^{=d} \rangle \leqslant \|f^{=d}\|_q \|f\|_{q'} \leqslant (Cq)^{d/2} \|f^{=d}\|_2 \cdot \alpha^{1-1/q}.$$

After rearranging we obtain

$$\|f^{=d}\|_2^2 \leqslant (e^2 Cq)^d \alpha^2.$$

The theorem follows by plugging in the value of q. □

Theorem 4.5 together with Lemma 3.15 also shows that Theorem 3.10 implies Theorem 1.5.

Combining Lemma 4.4 with Theorem 4.5, we have the following.

Lemma 4.6. *For each $c, C > 0$ there exist $c', n_0 > 0$ such that the following holds. Let $n > n_0$ and let G be an n-graded (C, c)-good group. Let $A \subseteq G$ and suppose that*

$$\alpha := \mu_G(A) \in (e^{-c'\sqrt{n}}, c').$$

Write $f = \frac{1_A}{\alpha}$ and $t = \frac{\log(1/\alpha)}{2}$. Then for all $1 \leqslant d \leqslant t$, we have

$$\|T_f\|_{V_{=d}} \leqslant \left(\frac{C' \log(1/\alpha)}{n} \right)^{d/2},$$

where $C' := \frac{10C}{c}$. Moreover,

$$\|T_f\|_{V_{>t}} \leqslant \left(\frac{\alpha}{n} \right)^{10}.$$

Proof. The lemma follows by applying Lemma 4.4 with the values of Q_d and Q which are promised by the goodness of G, and then bounding $\|f^{=d}\|_2$ using Theorem 4.5.

Let us begin with the range $d \leqslant t$. Since we know that G is c-strongly-quasirandom, he have that G is $((Q_d)_{d=1}^{\lceil n/2 \rceil - 1}, Q)$ graded where $Q_d \geqslant \left(\frac{cn}{d}\right)^d$ for all $d \leqslant t$. Hence, by Lemma 4.4, we have

$$\begin{aligned}
\|T_f\|_{V_{=d}} &\leqslant \frac{\|f^{=d}\|_2}{\sqrt{Q_d}} \\
&\leqslant \left\|f^{=d}\right\|_2 \cdot \left(\frac{cn}{d}\right)^{-d/2} \\
&= \frac{\|\alpha \cdot f^{=d}\|_2}{\alpha} \cdot \left(\frac{cn}{d}\right)^{-d/2} \\
&\leqslant \left(\frac{d}{cn}\right)^{d/2} \cdot \left(\frac{10C}{d}\right)^{d/2} \log(1/\alpha)^{d/2}
\end{aligned}$$

(using Lemma 4.5)

$$\leqslant \left(\frac{10C \log(1/\alpha)}{cn} \right)^{d/2},$$

which is the desired bound.

Next, consider the range $d \geqslant t$. For d in this range, and provided that n_0 is sufficiently large, we have from c-strong-quasirandomness that $Q_d \geqslant (\frac{cn}{t})^t$. By Lemma 4.4 we therefore have, by a similar computation to before,

$$\|T_f\|_{V_{>t}} \leqslant \|f\|_2 \left(\frac{cn}{t}\right)^{-t/2} \leqslant n^{-t/4} e^{-21t} \alpha^{-1/2} \leqslant \left(\frac{\alpha}{n}\right)^{10},$$

where we used

$$\frac{cn}{t} \geqslant e^{42} n^{1/2},$$

which holds provided c' is sufficiently small. □

Lemma 4.7. *Let $c, C, c' > 0$. Let G be an n-graded c-strongly-quasirandom group. Let $A \subseteq G$ and suppose that*

$$\alpha := \mu_G(A) \geqslant c'.$$

Write $f = \frac{1_A}{\alpha}$. Then for $1 \leqslant d < cn/(1+c)$ we have

$$\|T_f\|_{V_{=d}} \leqslant (c')^{-1/2} \left(\frac{d}{cn}\right)^{d/2},$$

and for $d \geqslant cn/(1+c)$ we have

$$\|T_f\|_{V_{=d}} \leqslant (c')^{-1/2} (1+c)^{-cn/(2(1+c))}.$$

Proof. The lemma follows by applying Lemma 4.4 with the values of Q_d and Q that are guaranteed by the goodness of G, and then upper-bounding $\|f^{=d}\|_2$ using $\|f^{=d}\|_2 \leqslant \|f\|_2 = \alpha^{-1/2} \leqslant (c')^{-1/2}$.

□

Lemmas 4.6, 4.7 and 3.15 together show that Theorem 3.10 implies Theorem 1.9.

The following is a version of Theorem 4.5 that is perhaps easier to comprehend.

Theorem 4.8. *For each $c, C > 0$ there exist $c', C' > 0$ such that the following holds. Let G be an n-graded, (c, C)-good group, let $f \colon G \to \{0, 1\}$, and suppose that $\alpha := \mathbb{E}[f] \geqslant e^{-c'\sqrt{n}}$. Then for all $d \in \mathbb{N}$, we have*

$$\|f^{=d}\|_2^2 \leqslant (C')^d \alpha^2 \log^d(e/\alpha). \tag{2}$$

Proof. Provided that C' is sufficiently large with respect to c' we may assume that $\alpha \leqslant c'$. Indeed, for $\alpha \geqslant c'$ the right hand side is greater than one, and we always have the trivial upper bound $\|f^{\leqslant d}\|_2^2 \leqslant \|f\|_2^2 = \alpha$. Let $r = c\sqrt{n}$, and note that in the case where $d > r$ the upper bound is trivial, as in this case the right-hand side of 2 is (again) greater than one.

Let $q = \log(1/\alpha)$. Let q' be the Hölder conjugate of q. For $d \leqslant r$, we have by (r, C)-hypercontractivity and Lemma 3.7 that

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

After rearranging we obtain

$$\|f^{=d}\|_2^2 \leqslant e^2(Cq)^d\alpha^2 = e^2C^d\alpha^2\log^d(1/\alpha).$$

This gives the right bound, provided C' is sufficiently large depending on C.

□

4.2 Upper bounds on the measures of product-free sets

Let us now show how Lemma 4.6 implies an upper bound on the measure of a product-free set in a good group.

Theorem 4.9. *There exist $c', n_0 > 0$ such that the following holds. Let $n > n_0$ and let G be a (c, C)-good n-graded group. Then every measurable product-free set in G has Haar measure at most $e^{-c'n^{1/3}}$.*

Before proving Theorem 4.9, let us note that together with Theorem 3.10 it implies Theorem 1.2. This follows since, by Lemma 3.15, for all $n > n_0$ every compact connected Lie group of min-rank n is a (c, C)-good n-graded group for some absolute constants $c, C > 0$.

Proof. Let $A \subseteq G$ be product-free and measurable; write $\alpha = \mathbb{E}[1_A]$ and $t = \frac{\log(1/\alpha)}{2}$. Assume w.l.o.g. that $\alpha < c'$, where c' is to be chosen later (if not, then replace A by a smaller product-free set). Let $f = \frac{1_A}{\alpha}$. Suppose for a contradiction that $\alpha \geqslant e^{-c'n^{1/3}}$. We have

$$0 = \langle T_f f, f\rangle = \mathbb{E}^3[f] + \sum_{d=1}^{\lfloor t\rfloor}\langle T_f f^{=d}, f^{=d}\rangle + \langle T_f f^{>t}, f^{>t}\rangle. \tag{3}$$

Provided c' is sufficiently small, we may now apply Lemma 4.6 and Theorem 4.5 with $1_A = f\alpha$ to obtain, for sufficiently large C' and all $1 \leqslant d \leqslant t$,

$$|\langle T_f f^{=d}, f^{=d}\rangle| \leqslant \|T_f\|_{V_{=d}}\|f^{=d}\|_2^2 \leqslant \left(\frac{C'\log^3(1/\alpha)}{nd^2}\right)^{d/2} \leqslant 100^{-d},$$

provided c' is sufficiently small depending on C'. We may also apply Lemma 4.6 to obtain

$$|\langle T_f f^{>t}, f^{>t}\rangle| \leqslant \|T_f\|_{V_{>t}}\|f\|_2^2 \leqslant \left(\frac{\alpha}{n}\right)^{10}\alpha^{-1}.$$

As $\mathbb{E}[f] = 1$, these two upper bounds contradict (3). □

4.3 Product mixing

The proof of Theorem 4.9 in fact gives the following stronger statement, which implies Theorem 1.11.

Theorem 4.10. *For any $\varepsilon > 0$, there exist c', n_0, such that the following holds. Let $n > n_0$ and let G be a (c, C)-good n-graded group. Let A, B, C be measurable subsets of G, each with Haar measure at least $e^{-c'n^{1/3}}$. Let $f = \frac{1_A}{\mu(A)}, g = \frac{1_B}{\mu(B)}, h = \frac{1_C}{\mu(C)}$. Then*

$$|\langle f * g, h\rangle - 1| < \varepsilon.$$

Proof. Assume without loss of generality that B has the smallest measure of the three sets. (Note that, while the trilinear form $\mathcal{T}(f, g, h) := \langle f * g, h\rangle$ is not quite symmetric with respect to permuting f, g and h, we may swap the positions of f and g or of g and h if we replace some of A, B and C by their inverses, meaning $A^{-1} := \{x^{-1} : x \in A\}$ etc, which have the same measures. So there is indeed no loss of generality in assuming the above.) Write $\mu(A) = \alpha, \mu(B) = \beta, \mu(C) = \gamma$ and $C' = \frac{C}{10c}$.

Note that $\langle f * g, h\rangle = \langle T_g f, h\rangle$. First we quickly handle the case where $\beta \geqslant c'$. Here we may apply Lemma 4.7 to obtain that $\|T_g - I_0\|_{2\to 2} < \varepsilon c'$, where I_0 denotes operator $F \mapsto \mathbb{E}[F]$ which sends a function to the constant function of the same expectation, provided that n_0 is sufficiently large. Using the fact that $\mathbb{E}[g] = 1$ we then have

$$|\langle T_g f, h\rangle - 1| = |\langle T_g f, h\rangle - \langle I_0 f, h\rangle| \leqslant \|T_g - I_0\|_{2\to 2} \cdot \|f\|_2 \cdot \|h\|_2 < \varepsilon \frac{c'}{\alpha^{1/2}\gamma^{1/2}} \leqslant \varepsilon,$$

yielding the conclusion of the theorem.

The proof of the case $\min\{\alpha, \beta, \gamma\} = \beta < c'$ proceeds similarly to the proof of Theorem 4.9. Let $t = \frac{\log(1/\beta)}{2}$. We have

$$\langle T_g f, h\rangle - 1 = \langle T_g f, h\rangle - \langle T_g f^{=0}, h^{=0}\rangle = \sum_{d=1}^{\lfloor t\rfloor} \langle T_g f^{=d}, h^{=d}\rangle + \langle T_g f^{>t}, h^{>t}\rangle. \tag{4}$$

Using Lemma 4.6, we obtain the upper bound

$$|\langle T_g f^{>t}, h^{>t}\rangle| \leqslant \|T_g\|_{V^{>t}} \|f\|_2 \|h\|_2 \leqslant \frac{\beta^{10}}{\alpha^{1/2} \cdot \gamma^{1/2} \cdot n^{10}} < \varepsilon/2, \tag{5}$$

provided that n_0 is sufficiently large. For $1 \leqslant d \leqslant t$, we use the upper bound

$$|\langle T_g f^{=d}, h^{=d}\rangle| \leqslant \|T_g\|_{V_{=d}} \|f^{=d}\|_2 \|h^{=d}\|_2. \tag{6}$$

By Lemma 4.6, we have $\|T_g\|_{V_{=d}} \leqslant \left(\frac{C' \log(1/\alpha)}{n}\right)^{d/2}$. By Theorem 4.8, we have the upper bound

$$\|f^{=d}\|_2^2 \leqslant C'^d \log^d(e/\alpha),$$

and similarly for h,

$$\|h^{=d}\|_2^2 \leqslant C'^d \log^d(e/\gamma).$$

Substituting the last three bounds into (6), we obtain

$$|\langle T_g f^{=d}, h^{=d}\rangle| \leqslant \left(\frac{C'^3 \log(e/\alpha) \log(e/\beta) \log(e/\gamma)}{n}\right)^{d/2} \leqslant \varepsilon 4^{-d}, \tag{7}$$

provided that c' is sufficiently small depending on C, c and ε.

The sum of the contribution from (5) and of those from (7) for $1 \leqslant d \leqslant t$, to the right-hand side of (4), is clearly less than ε, yielding $|\langle T_g f, h\rangle - 1| < \varepsilon$, as required. □

4.4 Equidistribution of convolutions

Let $A \subseteq G$ be a positive-measure subset of a good Lie group, and suppose that X is a G-homogeneous topological space (equipped with its G-invariant Haar probability measure μ_X), and that $B \subseteq X$ is a positive-measure subset. The next theorem states that as long as the measures of A and of B are not too small, applying a uniformly random element of A to a uniformly random element of B yields an almost uniformly random element of X (meaning, a random element with respect to the Haar probability measure). Note that Theorem 4.11 below, together with Lemma 3.15, imply Theorems 1.10 and 1.12.

Theorem 4.11. *For each $C, c, \varepsilon > 0$ there exists $c', n_0 > 0$, such that the following holds. Let $n > n_0$, let G be an n-graded (C, c)-good compact connected Lie group, and let X be a G-homogeneous topological space (equipped with the G-action $(g, x) \mapsto gx$), and let μ_X denote the G-invariant Haar probability measure on X. Suppose that $A \subseteq G$ and $B \subseteq X$ are measurable sets of Haar probability measures $\geqslant e^{-c'\sqrt{n}}$. Let μ_A denote the Haar probability measure on G, conditioned on the event A, and let μ_B denote the Haar probability measure on X conditioned on the event B, i.e. $\mu_B(Y) = \mu_X(B \cap Y)/\mu_X(B)$ for a measurable set $Y \subseteq X$, and $\mu_A(Z) = \mu_G(A \cap Z)/\mu_G(A)$ for a measurable set $Z \subseteq G$. Then the total variation distance between μ_X and the distribution of ab where $a \sim \mu_A$ and $b \sim \mu_B$ independently, is less than ε.*

Proof. Consider first the case where $X = G$ and G acts on itself by left multiplication; in this case, since the distribution of ab is $\mu_A * \mu_B$, we need to show that $\|\mu_A * \mu_B - \mu_G\|_{\mathrm{TV}} < \varepsilon$. We associate μ_A with the function $f_A = \frac{1_A}{\mu(A)} \in L^2(G)$ and similarly, we associate μ_B with the function $f_B = 1_B/\mu(B) \in L^2(G)$; it follows easily from the Cauchy-Schwarz inequality that $\|\mu_A * \mu_B - \mu_G\|_{\mathrm{TV}} \leqslant \|f_A * f_B - 1\|_2$. So our aim is now to prove that

$$\|f_A * f_B - 1\|_2 < \varepsilon. \tag{8}$$

In proving (8), we argue that we may assume, without loss of generality, that $\mu(B) \leqslant c'$. Indeed, if this does not hold, then write $B = \cup_{i\in I} B_i$ as a finite, disjoint union of sets B_i such that $c'/2 \leqslant \mu(B_i) \leqslant c'$. Once we have proved (8) for sets of measure at most c', we obtain the desired bound for B by convexity, noting that f_B is a convex combination of the functions f_{B_i}.

Set $t = \frac{\log(1/\mu(B))}{2}$. We now have

$$\|f_A * f_B - 1\|_2^2 = \sum_{d=1}^{t} \|T_{f_B} f_A^{=d}\|_2^2 + \|T_{f_B} f_A^{>t}\|_2^2.$$

Now

$$\|T_{f_B} f_A^{=d}\|_2 \leqslant \|T_{f_B}\|_{V_{=d}} \|f_A^{=d}\|_2.$$

and

$$\|T_{f_B} f_A^{\geqslant t}\|_2 \leqslant \|T_{f_B}\|_{V_{>t}} \|f_A^{>t}\|_2.$$

The bound (8) now easily follows from Theorem 4.8 and Lemma 4.6, similarly to in the proof of Theorem 4.10. Indeed, these yield

$$\|T_{f_B}\|_{V_{=d}} \|f_A^{=d}\|_2 \leqslant \left(\frac{C'^2 \log(1/\mu(A)) \log(1/\mu(B))}{n}\right)^{d/2} \leqslant \varepsilon 4^{-d}$$

for all $1 \leqslant d \leqslant t$, and

$$\|T_{f_B}\|_{V_{>t}} \|f_A^{>t}\|_2 \leqslant \left(\frac{\beta}{n}\right)^{10} \alpha^{-1/2} \leqslant \varepsilon/2,$$

provided that n_0 is sufficiently large and c' sufficiently small depending on c, C and ε.

To prove the general case, note that we may choose an arbitrary $x_0 \in X$ and set $\tilde{B} = \{b \in G : bx_0 \in B\}$. If $\tilde{b} \sim \mu_{\tilde{B}}$ then $\tilde{b}x_0 \sim \mu_B$, and if $g \sim \mu_G$ then $gx_0 \sim \mu_X$; it follows that $\|\mu_{ab} - \mu_X\|_{\mathrm{TV}} \leqslant \|\mu_A * \mu_{\tilde{B}} - \mu_G\|_{\mathrm{TV}}$, where μ_{ab} denotes the distribution of ab for $a \sim \mu_A$ and $b \sim \mu_B$ (independently). The result for the pair $(A, \tilde{B})$ therefore implies the result for (A, B). □

5 Growth in fine groups

In this section we show that Theorems 1.4, 1.7 and 1.8 follow from Theorem 3.14.

First, the theorem below is a variation on Theorem 4.5, with hypercontractivity replaced by weak hypercontractivity. Together with Lemma 3.16 it shows that Theorem 3.14 implies Theorem 1.4.

Theorem 5.1. *Let $c > 0$, $C > 1$ and let G be an (r, C)-weakly hypercontractive group. Let $f : G \to \{0, 1\}$ be measurable, and write $\alpha := \mathbb{E}[f]$. Let $d \in \mathbb{N}$ be such that $0 < d \leqslant \min\left(\frac{\log(1/\alpha)}{2}, r\right)$. Then*

$$\|f^{=d}\|_2^2 \leqslant \left(\frac{e \log(1/\alpha)}{d}\right)^{2Cd} \alpha^2.$$

Proof. Let $q = \log(1/\alpha)/d$, and let q' be the Hölder conjugate of q. Then by Hölder, weak hypercontractivity and Lemma 3.12, we have

$$\|f^{=d}\|_2^2 = \langle f, f^{=d} \rangle \leqslant \|f^{=d}\|_q \|f\|_{q'} \leqslant q^{Cd} \alpha^{1-1/q} \|f^{=d}\|_2.$$

The theorem follows by rearranging and substituting $\alpha^{-1/q} = e^d$. (Note that $C > 1$, by the definition of weak hypercontractivity.) □

The following lemma is a variant of Lemma 4.6 for fine groups – the proof is the same as for Lemma 4.6, only using Theorem 5.1 instead of Theorem 4.5. We will make use of it when proving our diameter bounds for fine groups.

Lemma 5.2. *For each $c > 0$ and $C > 1$ there exist $c', C', n_0 > 0$ such that the following holds. Let $n > n_0$ and let G be an n-graded (C, c)-fine group. Let $A \subseteq G$ be measurable, and suppose that*

$$\alpha := \mu_G(A) \in (e^{-c'n}, c').$$

Write $f = \frac{1_A}{\alpha}$ and $t = \frac{\log(1/\alpha)}{2}$. Then for $1 \leqslant d \leqslant t$ we have

$$\|T_f\|_{V_{=d}} \leqslant \left(\frac{e\log(1/\alpha)}{d}\right)^{Cd} \cdot \left(\frac{d}{cn}\right)^{d/2}.$$

We also have

$$\|T_f\|_{V_{>t}} \leqslant \frac{\alpha^{10}}{n^{10}}.$$

We now show that if f has small expectation, then most of the Fourier mass of f lies on the high degrees.

Lemma 5.3. *For each $c > 0$ and $C > 1$ there exist $c', n_0 > 0$ such that the following holds. Let $n > n_0$, let G be a (c, C)-fine n-graded group, and let $f\colon G \to \{0, 1\}$ be measurable. Suppose that $\alpha := \mathbb{E}[f] \geqslant e^{-c'n}$, and let $0 \leqslant t \leqslant \frac{\log(1/\alpha)}{10^6 C^2}$. Then $\|f^{\leqslant t}\|_2^2 \leqslant 2 \cdot 2\alpha^{1.99}$.*

Proof. Inserting a factor of 2 into the bound from Theorem 5.1, we have for any $0 < d \leqslant t$ that

$$\|f^{=d}\|_2^2 \leqslant 2^{-d}\left(\frac{2e\log(1/\alpha)}{d}\right)^{2Cd}\alpha^2,$$

and for $d = 0$ we have $\|f^{=0}\|_2^2 = \alpha^2$. Therefore $\|f^{\leqslant t}\|$ is upper-bounded by $2\alpha^2$ multiplied by the maximum of $\left(\frac{2e\log(1/\alpha)}{d}\right)^{2Cd}$ in the range where $d \leqslant \frac{\log(1/\alpha)}{10^6 C^2}$. Taking logs and computing the derivative with respect to d, it is easy to show that the maximum is obtained at the end point when $d = \frac{\log(1/\alpha)}{10^6 C^2}$. This shows that

$$\|f^{\leqslant t}\|_2^2 \leqslant 2\alpha^2 (2 \cdot 10^6 \cdot eC^2)^{\frac{2\log(1/\alpha)}{10^6 C}} \leqslant 2\alpha^{1.99},$$

as required. □

For smaller values of d, we have better bounds.

Lemma 5.4. *For each $\varepsilon, c > 0$ and $C > 1$, there exist $c', n_0 > 0$ such that the following holds. Let $n > n_0$, let G be a (C, c)-fine n-graded group, and let $f\colon G \to \{0, 1\}$ be measurable. Suppose that $\alpha := \mathbb{E}[f] \geqslant e^{-c'n}$ and let $d \in \mathbb{N}$ with $0 < d \leqslant \frac{\log(1/\alpha)}{n^{\varepsilon}}$. Then*

$$\|f^{\leqslant d}\|_2^2 \leqslant \alpha^{2 - n^{-\varepsilon/2}}.$$

Proof. Let c' be sufficiently small and n_0 sufficiently large with respect to ε, c, C. Let $t = \frac{\log(1/\alpha)}{n^\varepsilon}$. Similarly to in the proof of Lemma 5.3, it is easy to see that

$$\|f^{=d}\|_2^2 \leqslant \left(\frac{e\log(1/\alpha)}{t}\right)^{2Ct} \alpha^2 = (e \cdot n^\varepsilon)^{2C\log(1/\alpha)/n^\varepsilon}.$$

Now

$$(e \cdot n^\varepsilon)^{2C\log(1/\alpha)/n^\varepsilon} = \alpha^{-n^{-\varepsilon}\cdot 2C\log(en^\varepsilon)} \leqslant \alpha^{-n^{-\varepsilon/2}},$$

provided that $n \geqslant n_0$. □

5.1 Product mixing in fine groups

The next lemma is a version of Theorem 4.10, except for fine instead of for good groups. The only difference in the proof is that we apply Lemma 5.2 in place of Lemma 4.6 and Theorem 5.1 in place of Theorem 4.5.

Lemma 5.5. *For each $\varepsilon, c > 0$ and $C > 1$, there exists $\delta > 0$ such that the following holds. Let G be a (C, c)-fine, n-graded group, let $A, B, C \subseteq G$ be measurable sets of measures at least e^{-n^δ}, and let $f = \frac{1_A}{\mu(A)}, g = \frac{1_B}{\mu(B)}, h = \frac{1_C}{\mu(C)}$. Then $|\langle f * g, h\rangle - 1| < \varepsilon$.*

We remark that Lemma 5.5 is only weaker than the corresponding Theorem 4.10 for good groups. We include it even though the groups of interest to us are both good and fine. We decided to include the lemma mainly for aesthetic reasons. Our diameter bounds rely on Lemma 5.5 and we preferred to show that our diameter bounds hold for fine groups rather than groups that are both good and fine.

Below we use a trick of Nikolov and Pyber [33] (who observed that product mixing implies an upper bound on the diameter), to bound the diameter in fine groups.

Corollary 5.6. *For each $c > 0$ and $C > 1$, there exists $\delta > 0$ such that if G is a (C, c)-fine group and $\mathcal{A} \subseteq G$ is a measurable set of measure at least e^{-n^δ}, then $\mu(\mathcal{A}^2) > 1 - e^{-n^\delta}$, and $\mathcal{A}^3 = G$.*

Proof. The claim about $\mu(\mathcal{A}^2)$ follows by applying Lemma 5.5 while taking $A = B = \mathcal{A}$, $C = G \setminus \mathcal{A}^2$ and $\varepsilon = 1/2$ (in fact, any value of ε less than one, will do). As for the claim about $\mathcal{A}^3$, suppose for a contradiction that $\mathcal{A}^3 \neq G$. Let $x \in G \setminus \mathcal{A}^3$. Then $\mathcal{A}^2 \cap x\mathcal{A}^{-1} = \varnothing$. This contradicts Lemma 5.5 when setting $A = B = \mathcal{A}$ and $C = x\mathcal{A}^{-1}$. □

5.2 Non-Abelian Brunn–Minkowski for fine groups

The following theorem is a restatement of Theorem 1.8.

Lemma 5.7. *There exists absolute constants $c', n_0 > 0$ such that the following holds. Let G be a compact connected Lie group with $n := n(G) > n_0$, and let $A \subseteq G$ be a measurable set of measure at least $e^{-c'n}$. Then $\mu(A)^2 \geqslant \mu(A)^{0.1}$.*

Proof. First note that $\mu(A) > \frac{1}{2}$ implies that $A^2 = G$, as if $x \in G \setminus A^2$, then A and xA^{-1} are disjoint sets each of measure greater than $1/2$, a contradiction. By Corollary 5.6, we may also assume that $\mu(A) \leqslant e^{-n^{c'}}$, provided c' is sufficiently small.

Let $f = \frac{1_A}{\mu(A)}$ and $g = 1_{A^2}$. Then we have $\langle f * f, g\rangle = 1$. On the other hand by Cauchy–Schwarz we have

$$|\langle f * f, g\rangle| \leqslant \|f * f\|_2 \|g\|_2 = \|f * f\|_2 \sqrt{\mu(A^2)}.$$

This yields $\mu(A^2) \geqslant \frac{1}{\|f*f\|_2^2}$. Let $t = \frac{\log(1/\mu(A))}{10^6 C^2}$. We have $\|f*f\|_2^2 = \|f^{\leqslant t} * f^{\leqslant t}\|_2^2 + \|f^{\geqslant t} * f^{\geqslant t}\|_2^2$. By applying lemma 5.3 with 1_A we obtain

$$\|f^{\leqslant t} * f^{\leqslant t}\|_2^2 \leqslant \|f^{\leqslant t}\|_2^4 \leqslant \alpha^{-0.02}.$$

By applying Lemma 5.2 we obtain

$$\|f^{>t} * f^{>t}\|_2^2 \leqslant \|T_f\|_{V_{>t}}^2 \|f\|_2^2 \leqslant \left(\frac{\alpha}{n}\right)^{20} \alpha^{-1} \leqslant 1.$$

Combining the bounds completes the proof. □

Lemma 5.8. *For each $\varepsilon, c > 0$ and $C > 1$ there exist $\delta, n_0 > 0$ such that the following holds. Let $n > n_0$ and let G be a (C, c)-fine n-graded group. If*

$$\mu(A) := e^{-n^{\zeta}} \in \left(e^{-n^{1-\varepsilon}}, e^{-n^{\varepsilon}}\right),$$

then $\mu(A^2) \geqslant e^{-n^{\zeta-\delta}}$.

Proof. We may and shall assume, throughout the proof, that δ is sufficiently small depending on ε, C and c, and that n_0 is sufficiently large depending on δ. Let $f = \frac{1_A}{\mu(A)}$. As in the proof of the previous lemma, we have

$$\mu(A^2) \geqslant \frac{1}{\|f * f\|_2^2}. \tag{9}$$

Let $t_1 = \frac{\log(1/\mu(A))}{n^{4\delta}}$ and $t_2 = \frac{\log(1/\mu(A))}{10^6 C^2}$. We bound $\|f * f\|_2^2$ from above by decomposing it as follows:

$$\|f * f\|_2^2 = \|f^{<t_1} * f^{<t_1}\|_2^2 + \sum_{d=t_1}^{t_2} \|f^{=d} * f^{=d}\|_2^2 + \|f^{>t_2} * f^{>t_2}\|_2^2. \tag{10}$$

By applying Lemma 5.2, we obtain

$$\|f^{>t_2} * f^{>t_2}\|_2 \leqslant \|T_f\|_{V_{>t_2}} \|f\|_2 \leqslant \frac{\alpha^{10}}{n^{10}} \alpha^{-1/2} \leqslant 1.$$

Applying Lemma 5.4 (with αf in place of f, and ε taken to be 4δ), we have

$$\|f^{<t_1} * f^{<t_1}\|_2 \leqslant \|f^{<t_1}\|_2^2 \leqslant \alpha^{-2} \cdot \alpha^{2-n^{-2\delta}} \leqslant \frac{1}{4}\alpha^{-n^{-\delta}},$$

provided that δ is sufficiently small and n is sufficiently large depending on δ.

Finally, for $t_1 < d < t_2$ we combine Lemma 5.2 with Theorem 5.1 to obtain the upper bound

$$\|f^{=d} * f^{=d}\|_2 \leqslant \|Tf\|_{V_{=d}} \|f^{=d}\|_2 \leqslant \alpha^2 \cdot \left(\frac{e\log(1/\alpha)}{d}\right)^{2Cd} \cdot \left(\frac{cn}{d}\right)^{-d/2}$$
$$\leqslant 1 \cdot \left(\frac{e\log(1/\alpha)}{d}\right)^{2Cd} \cdot \left(\frac{c\log(1/\alpha)n^{\varepsilon}}{d}\right)^{-d/2}.$$

We may now use the fact that

$$\left(\frac{e\log(1/\alpha)}{d}\right) \leqslant e \cdot n^{4\delta}$$

to obtain

$$\|f^{=d} * f^{=d}\|_2 \leqslant \left(\frac{e\log(1/\alpha)}{d}\right)^{2Cd} \cdot (cn^{\varepsilon})^{-d/2} \leqslant n^{9\delta Cd} n^{-\varepsilon d/2} \leqslant \frac{1}{2n},$$

provided δ is sufficiently small. Substituting all of these upper bounds into (10) yields

$$\|f * f\|_2^2 \leqslant 2 + \tfrac{1}{4}\alpha^{-n^{-\delta}} \leqslant \alpha^{-n^{-\delta}}$$

provided n_0 is sufficiently large, and substituting this into (9) completes the proof. □

5.3 Diameter bounds for fine groups

Theorem 5.9. *For each $\varepsilon, c > 0$ and $C > 1$, there exist $m, n_0 > 0$, such that the following holds. Let $n > n_0$ and suppose that G is a (C,c)-fine n-graded group. Let $A \subseteq G$ be a measurable set with $\mu(A) > e^{-n^{1-\varepsilon}}$. Then $A^m = G$.*

Proof. Note that we may assume ε is as small as we please (depending on c and C). First apply Lemma 5.8 repeatedly (N times, say), until A^{2^N} has measure $\geqslant e^{-n^{\varepsilon}}$. In other words, $\mu(A^{m_1}) \geqslant e^{-n^{\varepsilon}}$, where m_1 depends upon ε alone. We can now apply Corollary 5.6 to obtain $A^{3m_1} = G$, provided that ε is sufficiently small depending on c and C. □

6 The strong quasirandomness of the simply connected compact Lie groups

In this section we describe the degree decomposition of the n-graded simply connected simple compact Lie groups in terms of their irreducible subrepresentations. We also show that all of them are c-strongly-quasirandom, for some absolute constant $c > 0$. In addition, we introduce the notion of *comfortable d-juntas*. These will be important in our proofs.

One of the goals of this section is to show that each Peter-Weyl ideal $W_\rho \subseteq V_{=d}$ contains a comfortable d-junta. This is useful because any linear operator on $L^2(G)$ that commutes with the action of G from both sides, has each W_ρ as an eigenspace. The operators we use in Section 8 will have this commuting property, and so when computing their eigenvalues we can simply consider the action of the relevant operator on a comfortable d-junta.

6.1 The Peter-Weyl theorem

We now recall some classical facts from the representation theory of compact groups. The Peter-Weyl theorem states that if G is a compact group, equipped with its Haar probability measure, then $L^2(G)$ has the following decomposition as an orthogonal direct sum:

$$L^2(G) = \bigoplus_{\rho \in \hat{G}} W_\rho,$$

where $\hat{G}$ denotes a complete set of complex irreducible unitary representations of G (here, *complete* means having one irreducible representation from each equivalence class of irreducible representations), and W_ρ is the subspace of $L^2(G)$ spanned by functions of the form $g \mapsto u^t \rho(g) v$, for $u, v \in V$, where V is the vector space on which ρ acts. The latter functions are known as the *matrix coefficients* or the *matrix entries* of ρ. The subspaces W_ρ are two-sided ideals (meaning, they are closed under both left and right actions of G), and they are also topologically closed; in fact, they are precisely the minimal non-zero topologically closed two-sided ideals of $L^2(G)$, and they are therefore irreducible as $G \times G$-modules (the $G \times G$ action being defined in the obvious way, with the first factor acting on $L^2(G)$ from the left and the second from the right). We call them the *Peter-Weyl ideals* of $L^2(G)$, though this terminology is non-standard. The space W_ρ can be decomposed as a direct sum of $\dim(\rho)$ irreducible left-representations.

Since the Peter-Weyl ideals W_ρ are precisely the minimal closed two-sided ideals of $L^2(G)$, every closed two-sided ideal of $L^2(G)$ can be decomposed as a direct sum of some of the W_ρ. Let $d \in \mathbb{N} \cup \{0\}$; since $V_{=d}$ is a closed, two-sided ideal of $L^2(G)$, there exists a set $\mathcal{L}_d$ of irreducible representations of G such that

$$V_{=d} = \bigoplus_{\rho \in \mathcal{L}_d} W_\rho.$$

If $\rho \in \mathcal{L}_d$ for some integer $0 \leqslant d < n/2$, we say that the *level* of ρ is equal to d. (Note that, since the $V_{=d}$ are pairwise orthogonal, the sets $\mathcal{L}_d$ are pairwise disjoint.)

6.2 The special orthogonal group $\mathrm{SO}(n)$

Our goal is now to show that for the group $\mathrm{SO}(n)$ each $\rho \in \mathcal{L}_d$ has dimension at least $(\frac{cn}{d})^d$ for some absolute constant $c > 0$, for each $d < n/2$. We will also show that the other irreducible representations of $\mathrm{SO}(n)$ all have dimension at least exponential in n, i.e. at least $\exp(c'n)$ for some absolute constant $c' > 0$.

We briefly recall Weyl's construction of the irreducible representations of SO(n). For more detail on Weyl's construction, the reader is referred for example to the book [12] of Fulton and Harris. (We note that, though the description in [12] is of SO($n, \mathbb{C}$), the irreducible representations of SO(n) := SO($n, \mathbb{R}$) are in a dimension-preserving one-to-one correspondence with those of its complexification SO($n, \mathbb{C}$).) We start by describing the irreducible representations of O(n) := $O(n, \mathbb{R})$. Let $V = \mathbb{R}^n$ denote the standard representation of O(n), defined by $\rho_V(g)(v) = g \cdot v$ — meaning, multiplication of the matrix g with the column-vector v. (We note, for later, that the restriction of this representation to SO(n) is known as the standard representation of SO(n).) For a partition $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ of some non-negative integer, let $d = \sum_{i=1}^{\ell} \lambda_i$. Consider the group algebra of the symmetric group on d elements, $\mathbb{R}[S_d]$, with the standard basis $\{e_g : \ g \in S_d\}$, and with multiplication defined by $e_g e_h = e_{gh}$ for $g, h \in S_d$. (Where there is no risk of confusion, we will sometimes write g in place of e_g, as an element of $\mathbb{R}[S_d]$, as is usual practice.) Let T be the standard Young tableau of shape λ with the numbers $1, 2, \ldots, \lambda_1$ (in order) in the first row, the numbers $\lambda_1 + 1, \lambda_1 + 2, \ldots, \lambda_1 + \lambda_2$ (in order) in the second row, and so on. Also, let P be the subgroup of S_d stabilising each of the rows of T (as sets), let Q be the subgroup of S_d stabilising each of the columns of T (as sets), and let

$$c_\lambda = \left(\sum_{g \in P} e_g \right) \left(\sum_{g \in Q} \mathrm{sign}(g) e_g \right)$$

be the *Young symmetrizer* of λ corresponding to T. The group S_d acts on $V^{\otimes d}$ from the right, permuting the factors:

$$(v_1 \otimes v_2 \otimes \ldots \otimes v_d) g = v_{g(1)} \otimes v_{g(2)} \otimes \ldots \otimes v_{g(d)},$$

and, extending linearly, so does $\mathbb{R}[S_d]$.

We define the *Weyl module* $\mathbb{S}_\lambda(V) := V^{\otimes d} c_\lambda$. Clearly, $\mathbb{S}_\lambda(V)$ is a left O(n)-submodule of $V^{\otimes d}$. It is reducible in general. However, we can obtain an irreducible left O(n)-module by considering $\mathbb{S}_{[\lambda]}(V) := V^{[d]} c_\lambda$, where $V^{[d]}$ is defined to be the intersection of the kernels of all $\binom{d}{2}$ linear maps on $V^{\otimes d}$ of the form

$$v_1 \otimes v_2 \otimes \ldots \otimes v_d \mapsto \langle v_i, v_j \rangle v_1 \otimes v_2 \otimes \ldots \otimes v_{i-1} \otimes v_{i+1} \otimes \ldots \otimes v_{j-1} \otimes v_{j+1} \otimes \ldots \otimes v_d.$$

Such linear maps are called *contractions*. It turns out that when the sum of the lengths of the first two columns of the Young diagram of λ is greater than n, we have $S_{[\lambda]}(V) = \{0\}$. The other modules $\mathbb{S}_{[\lambda]}(V)$ (corresponding to those partitions λ such that the sum of the first two columns of the Young diagram of λ is at most n) form a complete set of pairwise inequivalent irreducible complex representations of O(n).

Weyl's construction for SO(n) requires only one additional ingredient. We say two partitions λ and μ are *associated* if the sum of the lengths of the first column of λ and the first column of μ is equal to n, and the ith column of λ has the same length as the ith column of μ for each $i > 1$. If λ and μ are a pair of distinct associated partitions, then $\mathbb{S}_{[\lambda]}(V)$

and $\mathbb{S}_{[\mu]}(V)$ restrict to isomorphic representations of $\mathrm{SO}(n)$. If λ is self-associated (which happens iff n is even and the first column of λ has length $n/2$), then $\mathbb{S}_{[\lambda]}(V)$ restricts to a direct sum of two isomorphic irreducible representations of $\mathrm{SO}(n)$; if λ is not self-associated, then $\mathbb{S}_{[\lambda]}(V)$ restricts to an irreducible representation of $\mathrm{SO}(n)$, and if λ' is the partition associated to λ, then $\mathbb{S}_{[\lambda']}(V)$ restricts to the same irreducible representation of $\mathrm{SO}(n)$. In the latter case, it is customary to choose (as the representative of its equivalence class), the partition with first column of length less than $n/2$. Note that, importantly for us, for any partition λ with $\sum_i \lambda_i < n/2$, $\mathbb{S}_{[\lambda]}(V)$ is irreducible as an $\mathrm{SO}(n)$-representation, as well as being irreducible as an $\mathrm{O}(n)$-representation, and moreoever, as λ ranges over partitions of integers less than $n/2$, the $S_{[\lambda]}(V)$ are pairwise inequivalent as $\mathrm{SO}(n)$-representations, as well as being pairwise inequivalent as $\mathrm{O}(n)$-representations.

An alternative definition of the level of a representation

The purpose of this section is to show that for a partition λ with $\sum_i \lambda_i < n/2$, the level of the irreducible representation $\mathbb{S}_{[\lambda]}(V)$ of $\mathrm{SO}(n)$ is equal to $\sum_i \lambda_i$.

For $0 \leqslant d < n/2$, define (as above) $\mathcal{L}_d := \{\rho \in \widehat{\mathrm{SO}(n)} : \ \rho \text{ has level } d\}$, and define $\tilde{\mathcal{L}}_d$ to be the set of irreducible representations of $\mathrm{SO}(n)$ (up to equivalence) that have the form $S_{[\lambda]}(V)$, where $\sum_i \lambda_i = d$. We wish to show that $\mathcal{L}_d = \tilde{\mathcal{L}}_d$ for all $0 \leqslant d < n/2$.

Lemma 6.1. *Let $V = \mathbb{R}^n$ be the standard representation of $\mathrm{SO}(n)$ and let $0 \leqslant d < n/2$. Then all irreducible $\mathrm{SO}(n)$-subrepresentations of $V^{\otimes d}$ are elements of $\cup_{i=0}^{d} \tilde{\mathcal{L}}_i$.*

Proof. We prove this lemma by induction on d. The case where $d = 0$ is trivial. Let $d \in \mathbb{N}$ and assume the statement of the lemma holds whenever d is replaced by some $d' < d$. Recall that, since $V^{\otimes d}$ can be expressed a sum of $V^{[d]}$ and some other modules all isomorphic to $V^{\otimes(d-2)}$, all the irreducible $\mathrm{SO}(n)$-subrepresentations of $V^{\otimes d}$ appear either as $\mathrm{SO}(n)$-subrepresentations of the module $V^{[d]}$ or as $\mathrm{SO}(n)$-subrepresentations of $V^{\otimes(d-2)}$. By the induction hypothesis, it therefore suffices to show that each irreducible $\mathrm{SO}(n)$-subrepresentation of $V^{[d]}$ is an element of $\tilde{\mathcal{L}}_i$ for some i. Let $c_{\lambda,T}$ be the Young symmetrizer corresponding to the Young tableau T (not necessarily the standard one) of shape λ. It is well-known that the $c_{\lambda,T}$ (as λ ranges over all partitions of d and T over all Young tableaus of shape λ) spans a subspace of $\mathbb{R}[S_d]$ containing the class functions; in particular, we may write $\mathrm{Id} \in \mathbb{R}[S_d]$ as a real linear combination of the $c_{\lambda,T}$'s. It follows that $V^{[d]}$ is a sum of left $\mathrm{O}(n)$-modules of the form $V^{[d]}c_{\lambda,T}$, and clearly $V^{[d]}c_{\lambda,T}$ is isomorphic to $V^{[d]}c_\lambda = \mathbb{S}_{[\lambda]}(V)$, as either a left $\mathrm{O}(n)$-module or a left $\mathrm{SO}(n)$-module. This completes the proof of the lemma. □

The following lemma implies that $\mathcal{L}_d = \tilde{\mathcal{L}}_d$ for all $0 \leqslant d < n/2$.

Lemma 6.2. *If $0 \leqslant d < n/2$, then*

$$V_{=d} = \bigoplus_{\rho \in \tilde{\mathcal{L}}_d} W_\rho.$$

Proof. We prove the lemma by induction on d. For $d = 0$, the statement is trivial. Suppose now that $d > 0$. Since the spaces W_ρ are pairwise orthogonal, the induction hypothesis reduces our task to showing that

$$V_{\leqslant d} = \bigoplus_{i=0}^{d} \bigoplus_{\rho \in \tilde{\mathcal{L}}_d} W_\rho.$$

Let us write $\tilde{V}_{\leqslant d} := \bigoplus_{i=0}^{d} \bigoplus_{\rho \in \tilde{\mathcal{L}}_d} W_\rho$. We now use Lemma 6.1, namely that all the irreducible subrepresentations of $V^{\otimes d}$ are elements of $\tilde{\mathcal{L}}_i$ for some $i \leqslant d$. The matrix coefficients of the representation $V^{\otimes d}$ include the entries of the matrix $X^{\otimes d}$, where $X \in \mathrm{SO}(n)$ is the input matrix. These are exactly the degree-d monomials in the entries of X. Decomposing $V^{\otimes d}$ into irreducible representations, we see that all the homogeneous degree-d polynomials belong to $\tilde{V}_{\leqslant d}$; using the induction hypothesis again, all polynomials of degree at most $d-1$ belong to $\tilde{V}_{\leqslant d-1} \subset \tilde{V}_{\leqslant d}$, and therefore $V_{\leqslant d} \subseteq \tilde{V}_{\leqslant d}$. The reverse inclusion ($\tilde{V}_{\leqslant d} \subseteq V_{\leqslant d}$) follows immediately from the fact that if $\sum_i \lambda_i = d$, then the matrix coefficients of $\mathbb{S}_{[\lambda]}(V)$ are homogeneous degree-d polynomials in the entries of the input matrix. □

We can now extend the notion of level to all the representations of $\mathrm{SO}(n)$.

Definition 6.3. *Let ρ be an irreducible representation of* $\mathrm{SO}(n)$ *and let λ be the corresponding partition, whose Young diagram has first column of length at most $n/2$. Then we define the* level *of ρ to be $\sum_i \lambda_i$.*

Comfortable d-juntas

We now digress a little and show that Weyl's construction implies that each Peter-Weyl ideal W_ρ contains a certain 'nice' function. This will be used later, in Section 8.

Our 'nice' functions are as follows.

Definition 6.4. *The* comfortable d-juntas *on* $\mathrm{SO}(n)$ *are the functions on* $\mathrm{SO}(n)$ *of the form*

$$X \mapsto \sum_{\sigma \in S_d} a_\sigma x_{1\sigma(1)} \cdots x_{d\sigma(d)}$$

for $a_\sigma \in \mathbb{R}$.

We remark that we use the term 'junta' here, by analogy with juntas in the theory of Boolean functions (on $\{0,1\}^n$), because functions of the above form depend only upon the upper $d \times d$ minor, though we stress that we will be interested in the case where d is polynomial in n (e.g. $d \sim \sqrt{n}$), rather than just the case of d fixed and n large.

Letting $e_1, \ldots, e_n$ be the standard orthonormal basis of $\mathbb{R}^n$, since $\langle e_i, e_j \rangle = 0$ for all $i, j \in [d]$ we have $e_1 \otimes \ldots \otimes e_d \in V^{[d]}$. Therefore $(e_1 \otimes \ldots \otimes e_d)c_\lambda \in S_{[\lambda]}(V)$, and thus the function P_λ in $L^2(\mathrm{SO}(n))$ defined by

$$P_\lambda(X) := \langle X((e_1 \otimes \ldots \otimes e_d)c_\lambda), e_1 \otimes \ldots \otimes e_d \rangle$$

is a matrix coefficient of $\mathbb{S}_{[\lambda]}(V)$. Moreover, the function P_λ is clearly a comfortable d-junta: writing $c_\lambda = \sum_{\sigma \in S_d} \varepsilon_\sigma \sigma$, where $\varepsilon_\sigma \in \{-1, 0, 1\}$ for each $\sigma \in S_d$, we have

$$P_\lambda(X) = \sum_{\sigma \in S_d} \varepsilon_\sigma \prod_{i=1}^{d} x_{i\sigma(i)}.$$

Moreover, we clearly have $P_\lambda(\mathrm{Id}) = 1$, so P_λ is a non-zero element of $L^2(\mathrm{SO}(n))$. We obtain the following conclusion, upon which we rely crucially in the sequel.

Fact 6.5. *Let $0 \leqslant d < n/2$. For each irreducible representation $\rho \in \mathcal{L}_d$ of $\mathrm{SO}(n)$, the Peter-Weyl ideal W_ρ contains a nonzero comfortable d-junta.*

6.3 Getting strong quasirandomness

The following lower bound on the dimension of an irreducible representation of $\mathrm{SO}(n)$ follows immediately from the analysis in [38] of Weyl's original dimension formulae [41]. (We note that our comfortable d-junta machinery could be used to easily obtain a slightly weaker lower bound of $\binom{\lfloor n/2 \rfloor}{d}$. We use such an argument later, when showing strong quasirandomness for $\mathrm{SU}(n)$.)

Lemma 6.6. *If ρ is an irreducible representation of $\mathrm{SO}(n)$ of level $d \leqslant n$, then*

$$\dim(\rho) \geqslant \frac{(n-d)^d}{d!}.$$

We also need the following lower bound, whose proof is deferred to the Appendix.

Lemma 6.7. *Let $n \geqslant 5$. If ρ is an irreducible representation of $\mathrm{SO}(n)$ of level $d \geqslant n/2$, then*

$$\dim(\rho) \geqslant \exp(n/32).$$

Lemmas 6.6 and 6.7 immediately give strong quasirandomness.

Theorem 6.8. *For each $n \geqslant 2$, the n-graded group $\mathrm{SO}(n)$ is c-strongly-quasirandom, for some absolute constant $c > 0$.*

6.4 The spin group Spin(n)

The strong quasirandomness of the group $\mathrm{Spin}(n)$ follows from the fact that it is a double cover of $\mathrm{SO}(n)$.

Theorem 6.9. *For each $n \geqslant 3$, the n-graded group $\mathrm{Spin}(n)$ is c-strongly-quasirandom, for some absolute constant $c > 0$.*

Proof. Recall that the spin group Spin(n) is the double-cover of SO(n) for all $n \geqslant 2$. It is a simply-connected real Lie group for all $n \geqslant 3$, so its complex irreducible representations are in an explicit (and dimension-preserving) one-to-one correspondence with those of its Lie algebra. It has the same Lie algebra as SO(n); this Lie algebra $\mathfrak{so}(n, \mathbb{R})$ is simple for all $n \geqslant 5$, and its complexification $\mathfrak{so}(n, \mathbb{C})$ is likewise simple (for all $n \geqslant 5$), so by e.g. [12] (26.14), for all $n \geqslant 5$ the complex irreducible representations of $\mathfrak{so}(n, \mathbb{R})$ are restrictions of unique complex irreducible representations of $\mathfrak{so}(n, \mathbb{C})$. The dimensions of the complex irreducible representations of Spin(n) (for all $n \geqslant 5$) are therefore given by equations (24.29) and (24.41) in [12] (pages 408 and 410). For all $n = 2k+1 \geqslant 5$ odd, we have

$$\dim(\rho_\lambda) = \prod_{1 \leqslant i < j \leqslant k} \frac{\lambda_i - \lambda_j - i + j}{j - i} \prod_{1 \leqslant i \leqslant j \leqslant k} \frac{\lambda_i + \lambda_j + 2k + 1 - i - j}{2k + 1 - i - j},$$

where the k-tuple $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots \lambda_k)$ ranges over all k-tuples defined by

$$\lambda_i = a_i + a_{i+1} + \ldots + a_{k-1} + \tfrac{1}{2} a_k$$

for some $(a_i)_{i=1}^k \in (\mathbb{N} \cup \{0\})^k$. The case of a_k even corresponds to irreducible representations of Spin(n) that are also irreducible representations of SO(n); the dimensions of these were bounded previously. The case of a_k odd corresponds to 'new' irreducible representations of Spin(n), but the above equation implies that any such has dimension at least $2^{\Omega(n)}$. This calculational check is deferred to the Appendix (see Section A.2). □

6.5 The compact symplectic group Sp(n)

We now turn to the case of the compact symplectic group. This group has two common descriptions, and both will be important for us.

We start with its description as the unitary group over the field of quarternions, $\mathbb{H}$. A shorthand is useful at this point: a *quaternionic matrix* is a matrix with quaternion entries. The conjugate $\bar{x}$ of a quaternion $x = a+ib+jc+kd$ is defined, as usual, by $\bar{x} = a-ib-jc-kd$. The conjugate $\bar{M}$ of a quaternionic matrix M is defined by $(\bar{M})_{s,t} = \overline{M_{s,t}}$ for all s, t, and the Hermitian conjugate M^* is defined by $M^* = (\bar{M})^t$. We say an n by n quaternionic matrix is *unitary* if

$$M^* M = I = M M^*;$$

we note that the second equality above is equivalent to the first. For $n \in \mathbb{N}$, the *quarternionic unitary group* $U_n(\mathbb{H})$ is defined to be the group of n by n quaternionic unitary matrices. This is one way of viewing Sp(n).

A second way is obtained as follows. Note that an n by n quaternionic matrix can be written in the form $A + jB$, with A and B being complex n by n matrices. If $A + jB$ and $C + jD$ are two quaternionic matrices, then their product satisfies

$$(A + jB)(C + jD) = AC - \bar{B}D + j(BC + \bar{A}D),$$

and the Hermitian conjugate of the quaternionic matrix $A + jB$ satisfies $(A + jB)^* = A^* - jB^t$. It follows that the map

$$\Phi : \mathbb{H}^{n\times n} \to \mathbb{C}^{2n\times 2n}$$

from n by n quaternionic matrices to $2n$ by $2n$ complex matrices defined by

$$\Phi(A + jB) = \begin{pmatrix} A & -\bar{B} \\ B & \bar{A} \end{pmatrix}$$

is an (injective) ring homomorphism that preserves Hermitian conjugates. It is easily checked that $M \in \mathbb{H}^{n\times n}$ is unitary if and only if $\Phi(M) \in \mathbb{C}^{2n\times 2n}$ is unitary (writing $M = A + jB$, either condition holds iff $A^*A + B^*B = I$ and $A^tB = B^tA$ both hold). Finally, defining

$$\Omega := \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

we observe that an arbitrary unitary matrix

$$X = \begin{pmatrix} A & C \\ B & D \end{pmatrix} \in \mathbb{C}^{2n\times 2n}$$

satisfies $X^t\Omega X = \Omega$ (or, equivalently, $X^t\Omega = \Omega X^*$) if and only if $C = -\bar{B}$ and $D = \bar{A}$, i.e. if and only if X lies in $\Phi(U_n(\mathbb{H}))$. It follows that

$$\Phi(U_n(\mathbb{H})) = \{X \in \mathbb{C}^{2n\times 2n} : \ X^t\Omega X = \Omega\} \cap U_{2n}(\mathbb{C}).$$

Hence, we can also view $\mathrm{Sp}(n)$ as

$$\{X \in \mathbb{C}^{2n\times 2n} : \ X^t\Omega X = \Omega\} \cap U_{2n}(\mathbb{C}),$$

i.e., as the intersection of the compact Lie group $U_{2n}(\mathbb{C})$ with

$$\mathrm{Sp}(2n, \mathbb{C}) := \{X \in \mathbb{C}^{2n\times 2n} : X^t\Omega X = \Omega\}.$$

(The group $\mathrm{Sp}(2n, \mathbb{C})$ is known as the *complex symplectic group*, and it is the complexification of $\mathrm{Sp}(n)$.) It can be checked (e.g. by taking the Pfaffian of the equation $X^t\Omega X = \Omega$) that any element of $\mathrm{Sp}(2n, \mathbb{C})$ has determinant one, so $U_{2n}(\mathbb{C})$ can be replaced by $\mathrm{SU}_{2n}(\mathbb{C}) = \mathrm{SU}(2n)$ in the above identification:

$$\mathrm{Sp}(n) = \{X \in \mathbb{C}^{2n\times 2n} : \ X^t\Omega X = \Omega\} \cap \mathrm{SU}(2n).$$

Weyl's construction in Sp(n)

Weyl's construction for $\mathrm{Sp}(n)$ is similar to his construction for $\mathrm{O}(n)$. Our exposition follows Fulton and Harris [12], as before.

The standard representation of $\mathrm{Sp}(n)$ is given by the second description above, regarding the elements of $\mathrm{Sp}(n)$ as $2n$ by $2n$ complex unitary matrices, and taking their natural (left) action on $\mathbb{C}^{2n}$.

Let $V = \mathbb{C}^{2n}$ denote the standard representation of $\mathrm{Sp}(n)$. Similarly to in the $\mathrm{O}(n)$ case, we have contraction maps $\psi_{i,j} \colon V^{\otimes d} \to V^{\otimes(d-2)}$, given by

$$\psi_{i,j} : \; v_1 \otimes \cdots \otimes v_d \mapsto Q(v_i, v_j) v_1 \otimes \cdots \hat{v}_i \otimes \cdots \otimes \hat{v}_j \otimes \cdots \otimes v_d,$$

where $\hat{v}$ denotes that v is omitted from the tensor product, and $Q(v, w) := v^t \Omega w$, where Ω is the matrix above.

Let $V^{\langle d \rangle}$ be the intersection of the kernels of all the contractions $\psi_{i,j}$; since the elements of $\mathrm{Sp}(n)$ preserve the skew-symmetric form Q, $V^{\langle d \rangle}$ is a left $\mathrm{Sp}(n)$-module. Moreover, $V^{\langle d \rangle}$ is acted upon by S_d from the right, by permutation of the factors. (The fact that this action preserves $V^{\langle d \rangle}$ follows from the skew-symmetry of Q.) For a partition $\lambda \vdash d$, we define $S_{\langle \lambda \rangle}(V)$ to be the representation (or left $\mathrm{Sp}(n)$-module) $V^{\langle d \rangle} c_\lambda$, where c_λ is the Young symmetrizer defined in Section 6.2. It turns out that $S_{\langle \lambda \rangle}(V)$ is nonzero precisely when (the Young diagram of) λ has at most n rows, and the nonzero left $\mathrm{Sp}(n)$-modules of this form constitute a complete set of pairwise inequivalent complex irreducible representations of $\mathrm{Sp}(n)$. Note that the situation here is, if anything, even simpler than that for $\mathrm{SO}(n)$, where we have to worry, if only momentarily, about distinct $S_{[\lambda]}(V)$ (for two distinct values of λ) restricting to the same irreducible representation of $\mathrm{SO}(n)$.

Recall that the *level* of an irreducible representation ρ of $\mathrm{Sp}(n)$ was defined to be the non-negative integer d such that $\rho \in \mathcal{L}_d$, where

$$V_{=d}^{\mathrm{Sp}(n)} = \bigoplus_{\rho \in \mathcal{L}_d} W_\rho.$$

Let λ be a partition such that $\sum_i \lambda_i < n/2$. Our next aim is to show that the level of the irreducible representation $S_{\langle \lambda \rangle}(V)$ is equal to $\sum_i \lambda_i$, as in the case of $\mathrm{SO}(n)$.

Earlier, in Section 3, we wrote our input quarternion matrix as $A + iB + jC + kD$, where A, B, C, D are n by n real matrices, and we defined $V_{\leqslant d}$ to consist of the degree $\leqslant d$ polynomials in the entries of the matrices A, B, C and D. Alternatively, we may write the input matrix as $A + jB$, where A and B are n by n complex matrices; then $V_{\leqslant d}$ may be viewed as the vector space of degree $\leqslant d$ polynomials in the entries of $A, B \in \mathbb{C}^{n \times n}$ and their complex conjugates. Finally, we may use the identification Φ above (in Section 6.5) of a quarternion matrix $A + jB$ with the $2n \times 2n$ complex matrix $\begin{pmatrix} A & -\bar{B} \\ B & \bar{A} \end{pmatrix}$ to obtain that $V_{\leqslant d}$ simply consists of polynomials of degree at most d in the entries of $\begin{pmatrix} A & -\bar{B} \\ B & \bar{A} \end{pmatrix}$; this is precisely what we need to generalise our $\mathrm{SO}(n)$ proofs.

For each $0 \leqslant d < n/2$, let $\tilde{\mathcal{L}}_d$ denote the set of irreducible representations of $\mathrm{Sp}(n)$ (up to equivalence) of the form $\mathbb{S}_{\langle \lambda \rangle}(V)$, where λ is a partition of d. Using the fact that, as with $\mathrm{SO}(n)$, any irreducible $\mathrm{Sp}(n)$-subrepresentation of $V^{\otimes d}$ appears either as an $\mathrm{Sp}(n)$-subrepresentation of the module $V^{\langle d \rangle}$ or as an $\mathrm{Sp}(n)$-subrepresentation of $V^{\otimes(d-2)}$, the proof of Lemma 6.1 generalizes straightforwardly to give:

Lemma 6.10. *Let $V = \mathbb{C}^{2n}$ be the standard representation of* $\mathrm{Sp}(n)$ *and let* $0 \leqslant d < n/2$. *Then all irreducible* $\mathrm{Sp}(n)$*-subrepresentations of* $V^{\otimes d}$ *are elements of* $\cup_{0 \leqslant i \leqslant d} \tilde{\mathcal{L}}_i$.

Then, taking the input matrix X in the form $\begin{pmatrix} A & -\bar{B} \\ B & \bar{A} \end{pmatrix}$, the proof of Lemma 6.2 generalises straightforwardly to give:

Lemma 6.11. *Let* $0 \leqslant d < n/2$. *Then* $V_{=d}^{\mathrm{Sp}(n)} = \bigoplus_{\rho \in \tilde{\mathcal{L}}_d} W_\rho$.

As in the $\mathrm{SO}(n)$ case, we may now extend the notion of level to all irreducible representations of $\mathrm{Sp}(n)$.

Definition 6.12. *Let ρ be an irreducible representation of* $\mathrm{Sp}(n)$. *The* level *of ρ is the integer d such that ρ is isomorphic to $S_{\langle\lambda\rangle}(V)$, where λ is a partition of d.*

Comfortable d-juntas

We say that a monomial is a comfortable d-junta if for some choice of $\sigma \in S_d$ and $q_1, \ldots, q_d \in \{\mathbf{i}, \mathbf{j}, \mathbf{k}, \mathbf{real}\}$ it is the product $\prod_{i=1}^d (x_{i,\sigma(i)})_{q_i\text{-part}}$. We say that a polynomial is a *comfortable d-junta* if it is a linear combination of comfortable monomials. The proof for $\mathrm{SO}(n)$ generalizes easily to give the following.

Lemma 6.13. *Let ρ be a representation of level $d \leqslant n$ of* $\mathrm{Sp}(n)$. *Then W_ρ contains a non-zero comfortable d-junta.*

Proof. Letting $e_1, \ldots, e_{2n}$ be the standard orthonormal basis of $\mathbb{C}^{2n}$, since $Q(e_i, e_j) = 0$ for all $i, j \in [n]$ we have $e_1 \otimes \ldots \otimes e_d \in V^{\langle d \rangle}$. Therefore $(e_1 \otimes \ldots \otimes e_d)c_\lambda \in S_{\langle\lambda\rangle}(V)$, and thus the function P_λ in $L^2(\mathrm{Sp}(n))$ defined by

$$P_\lambda(X) := \langle X((e_1 \otimes \ldots \otimes e_d)c_\lambda), e_1 \otimes \ldots \otimes e_d \rangle$$

is a matrix coefficient of $\mathbb{S}_{\langle\lambda\rangle}(V)$. Note that this is exactly the same function as we exhibited for $\mathrm{SO}(n)$, except that its domain is $\mathrm{Sp}(n)$ rather than $\mathrm{SO}(n)$. The function P_λ is clearly a comfortable d-junta: writing $c_\lambda = \sum_{\sigma \in S_d} \varepsilon_\sigma \sigma$, where $\varepsilon_\sigma \in \{-1, 0, 1\}$ for each $\sigma \in S_d$, we have

$$P_\lambda(X) = \sum_{\sigma \in S_d} \varepsilon_\sigma \prod_{i=1}^d x_{i,\sigma(i)}.$$

Writing $X = \begin{pmatrix} A & -\bar{B} \\ B & \bar{A} \end{pmatrix}$, we see that each $x_{i,\sigma(i)}$ appearing above is actually $a_{i,\sigma(i)}$ (since $d \leqslant n$ and $\sigma \in S_d$); writing each $a_{i,\sigma(i)}$ as a sum of its real and imaginary parts and expanding, we see that P_λ is indeed a sum of monomials of the required form (in fact, with each q_i being either **real** or **i**).

The function P_λ is non-zero element of $L^2(\mathrm{Sp}(n))$ since (as with $\mathrm{SO}(n)$) we have $P_\lambda(\mathrm{Id}) = 1$. □

c-strong-quasirandomness

The dimensions of the complex irreducible representations of $\mathrm{Sp}(n)$ are given by equation (24.19) in [12]. These representations are in one-to-one correspondence with partitions of non-negative integers, whose Young diagrams have at most n rows; the dimension of the irreducible representation ρ_λ corresponding to the partition $\lambda = (\lambda_1, \ldots, \lambda_n)$ (with $\lambda_1 \geqslant \ldots \geqslant \lambda_n$) is given by

$$\dim(\rho_\lambda) = \prod_{1 \leqslant i < j \leqslant n} \frac{\lambda_i - \lambda_j - i + j}{j - i} \prod_{1 \leqslant i \leqslant j \leqslant n} \frac{\lambda_i + \lambda_j + 2n + 2 - i - j}{2n + 2 - i - j}.$$

As with the (odd) special orthogonal groups, we define the *level* of the representation ρ_λ to be the number of cells in the Young diagram of λ (i.e., it is the non-negative integer of which λ is a partition). The proof of the following lemma is similar to that of Lemma 6.7. We defer it to the Appendix.

Lemma 6.14. *If ρ_λ is an irreducible representation of $\mathrm{Sp}(n)$ of level $d \geqslant n/2$, then $\dim(\rho_\lambda) \geqslant \exp(n/16)$.*

Again, as with the special orthogonal groups, the following lower bound is immediate from the analysis in [38] of Weyl's dimension formulae [41].

Lemma 6.15. *If ρ is an irreducible representation of $\mathrm{Sp}(n)$ of level $d \leqslant n$, then*

$$\dim(\rho) \geqslant \frac{(n-d)^d}{d!}.$$

These yield the following.

Theorem 6.16. *For each $n \geqslant 2$, the n-graded group $\mathrm{Sp}(n)$ is c-strongly-quasirandom, for some absolute constant $c > 0$.*

6.6 The special unitary group SU(n)

The final group to consider is the special unitary group. We start by relating our degree decomposition of $L^2(\mathrm{SU}(n))$ to the decomposition of $L^2(\mathrm{SU}(n))$ into Peter-Weyl ideals.

Degree decomposition

Earlier, we defined $V_{\leqslant d}$ to consist of the polynomials of (total) degree at most d polynomials in the real parts and the imaginary parts of the entries of the input matrix $X \in \mathrm{SU}(n)$. Equivalently, $V_{\leqslant d}$ consists of the polynomials of (total) degree at most d in the entries of the input matrix and their complex conjugates.

Weyl's construction for SU(n)

We now recall Weyl's construction of the irreducible representations of $\mathrm{SU}(n)$, and deduce from it the consequences we need. (As before, for more detail on Weyl's construction, the reader is referred to Fulton and Harris [12], noting that the complex irreducible representations of $\mathrm{SU}(n)$ are the same as those of $\mathrm{SL}(n,\mathbb{C})$, since $\mathrm{SU}(n)$ is a maximal compact subgroup of $\mathrm{SL}(n,\mathbb{C})$.) Let $V = \mathbb{C}^n$ denote the standard representation of $\mathrm{SU}(n)$, defined by $\rho_V(g)(v) = g \cdot v$. For a partition λ with at most $n-1$ parts, let $d = d(\lambda) = \sum_i \lambda_i$. Let T be the standard Young tableau of shape λ (defined in Section 6.2) and let c_λ be the corresponding Young symmetrizer (also defined in Section 6.2). We define the corresponding *Weyl module* by $\mathbb{S}_\lambda(V) := V^{\otimes d} c_\lambda$. Clearly, $\mathbb{S}_\lambda(V)$ is a left $\mathrm{SU}(n)$-submodule of $V^{\otimes d}$. The modules $\mathbb{S}_\lambda(V)$, as λ ranges over all partitions with at most $n-1$ parts, constitute a complete set of pairwise inequivalent complex irreducible representations of $\mathrm{SU}(n)$. (Unlike in the cases of $\mathrm{SO}(n)$ and $\mathrm{Sp}(n)$, we do not need to pass to a subrepresentation of the Weyl module; the latter is already irreducible as a left $\mathrm{SU}(n)$-module.)

Unlike in the case of $\mathrm{Sp}(n)$, however, the complex conjugates of the entries of the input matrix are no longer matrix coefficients of the standard representation. Instead, they are matrix coefficients of the dual of the standard representation, i.e. they are the entries of the matrix $(A^{-1})^t = \bar{A}$. (Recall that the dual of a representation ρ is the representation ρ^* defined by $\rho^*(g) = (\rho(g^{-1}))^t$.) We note that $(\mathbb{S}_\lambda(V))^* \cong \mathbb{S}_\lambda(V^*) = (V^*)^{\otimes d} c_\lambda$.

We have three notions of level for a representation ρ of $\mathrm{SU}(n)$, and our goal is to show that they agree when the level is $< n/2$.

Recall that we defined $V_{=0} := V_{\leqslant 0}$ (the space of constant functions), and $V_{=d} = V_{\leqslant d} \cap V_{\leqslant d-1}^{\perp}$ for each $d \in \mathbb{N}$. Since $V_{=d}$ is closed under both left and right actions of $\mathrm{SU}(n)$ for each $d \in \mathbb{N} \cup \{0\}$, there exists a set $\mathcal{L}_d$ of irreducible representations of $\mathrm{SU}(n)$ such that

$$V_{=d}^{\mathrm{SU}(n)} = \bigoplus_{\rho \in \mathcal{L}_d} W_\rho.$$

If $\rho \in \mathcal{L}_d$ for some $0 \leqslant d < n/2$, we define the *level* of ρ to be d. (Note that, since the $V_{=d}$ are pairwise orthogonal, the sets $\mathcal{L}_d$ are pairwise disjoint.)

In addition, we define the *tensor level* of an irreducible representation ρ of $\mathrm{SU}(n)$ to be the minimal non-negative integer d such that there exist $d_1, d_2 \in \mathbb{N} \cup \{0\}$ with $d_1 + d_2 = d$ and with ρ being isomorphic to a subrepresentation of $V^{\otimes d_1} \otimes V^{*\otimes d_2}$. We write $\bar{\mathcal{L}}_d$ for the set of irreducible representations of $\mathrm{SU}(n)$ (up to equivalence) that have tensor level equal to d.

The third notion of level will soon be described in terms of the Young diagram/partition corresponding to the representation.

The following analogue of Lemma 6.1 is immediate, and implies that for $0 \leqslant d < n/2$, having level d is equivalent to having tensor level d.

Lemma 6.17. *If $0 \leqslant d < n/2$, then $V_{=d}^{\mathrm{SU}(n)} = \bigoplus_{\rho \in \bar{\mathcal{L}}_d} W_\rho$.*

Step vectors

For an irreducible representation $\rho = \mathbb{S}_\lambda(V)$ of $\mathrm{SU}(n)$ with corresponding partition $\lambda = (\lambda_1, \dots, \lambda_{n-1})$, write $a_i := \lambda_i - \lambda_{i+1}$ for each $i \in [n-1]$. We call the vector $(a_1, \dots, a_{n-1})$ the *step vector* of the representation ρ (or, abusing terminology slightly, the *step vector* of the partition λ). We order such vectors with respect to the lexicographic ordering, i.e. $\lambda >_{\text{lex}} \lambda'$ iff $\lambda_i > \lambda'_i$ where $i = \min\{j : \lambda_j \neq \lambda'_j\}$. For a (not necessarily irreducible) representation ρ of $\mathrm{SU}(n)$, its step vector is defined to be the lexicographically largest step vector of an irreducible subrepresentation of ρ.

The step vector is better-behaved than the corresponding partition, with respect to taking duals and tensors. The dual of ρ has the reversed step vector $(a_{n-1}, \dots, a_1)$, corresponding to the partition $(a_1 + \dots + a_{n-1}, \cdots, a_1 + a_2, a_1)$. Moreover, if ρ_1 and ρ_2 are two representations whose step vectors are $(a_1, \dots, a_{n-1})$ and $(b_1, \dots, b_{n-1})$, then the step vector of their tensor product is $(a_1 + b_1, \dots, a_{n-1} + b_{n-1})$. [reference????]

We say that an irreducible representation $\rho = \mathbb{S}_{\lambda(V)}$ of $\mathrm{SU}(n)$ is *efficient* if $\lambda_{\lfloor \frac{n}{2} \rfloor} = 0$. Equivalently, ρ is efficient if its step vector $w = (a_1, \dots, a_{n-1})$ has the property that its *second half* $w'' := (a_{\lceil n/2 \rceil}, \dots, a_{n-1})$ consists of zeros. We say that ρ is *dually-efficient* if its *first half* $w' := (a_1, \dots, \dots, a_{\lceil n/2 \rceil - 1})$ consists of zeroes. Alternatively, in terms of the corresponding partition λ, the irreducible representation ρ is dually-efficient if $\lambda_1 = \lambda_2 = \dots = \lambda_{\lceil n/2 \rceil - 1}$. (Note that the dual to each efficient representation is dually-efficient, and if n is odd, the converse also holds. For n even, our definition leads to a somewhat arbitrary choice of how to handle the middle part of the step vector, but this does not matter when the level is smaller than $n/2$.) We call the partition α with step vector w' the *efficient truncation* of λ and the partition β with step vector w' the *dually-efficient truncation* of λ.

Definition 6.18. *We define the* total level *of a representation $\rho = \mathbb{S}_\lambda(V)$ with step vector $(a_1, \dots, a_{n-1})$ to be*

$$\sum_{i=1}^{n-1} a_i \min\{i, n-i\}.$$

For each $d \in \mathbb{N} \cup \{0\}$, let $\tilde{\mathcal{L}}_d$ denote set of irreducible representations of $\mathrm{SU}(n)$ with total level d. Our next aim is to show that $\mathcal{L}_d = \tilde{\mathcal{L}}_d$ for all $0 \leqslant d < n/2$. For this, we first need the following.

Lemma 6.19. *Let $\mathbb{S}_\lambda(V)$ be an irreducible representation of $\mathrm{SU}(n)$ of total level d, with λ having α as its efficient truncation and β as its dually-efficient truncation. Then the representation $\mathbb{S}_\alpha(V) \otimes \mathbb{S}_\beta(V)$ can be decomposed as a direct sum of one copy of the representation $\mathbb{S}_\lambda(V)$ and some other irreducible representations, all of which have total level less than d.*

Proof. This follows from the Littlewood-Richardson rule, e.g. as given in Fulton and Harris [12, Section A.8]. Indeed, by the Littlewood-Richardson rule, writing $\alpha = (\alpha_1, \dots, \alpha_\ell)$, the irreducible constituents of $\mathbb{S}_\beta(V) \otimes \mathbb{S}_\alpha(V)$ are exactly those $\mathbb{S}_\lambda(V)$ such that the Young diagram of λ can be produced by the following algorithm. Take the Young diagram of β, and first add α_1 new boxes to the rows (in such a way as to produce the Young diagram of another partition, but with no two of the α_1 added boxes being added to the same column), and place a '1' in each of these α_1 new boxes. Then add a further α_2 boxes to the rows (again in such a way as to produce the Young diagram of another partition, but with no two of the α_2 added boxes being added to the same column), and place a '2' in each of these α_2 new boxes. Continue in this way (so that at the last step, α_ℓ new boxes are added). Now consider the sequence of length $\alpha_1 + \ldots + \alpha_\ell$ formed by concatenating the reversed rows of newly-added boxes, and check that if one looks at the first t entries in this sequence (for any t between 1 and $\alpha_1 + \ldots + \alpha_\ell$), the integer p appears at least as many times as the integer $p+1$ among these first t entries, for any $1 \leqslant p < \ell$. If this 'concatenation' condition holds, keep the Young diagram / partition; if not, reject it.

It is easy to check that the only way of performing this algorithm in such a way as to obtain a partition of level at least d, is to produce the (Young diagram of the) partition λ itself: the first α_1 new boxes must all be added to the first row of the Young diagram of β, the second α_2 new boxes must all be added to the second row, and so on. Indeed, if at the jth stage (when adding α_j new boxes containing the integer j), any box is added to a row above the jth row, then (inductively) one sees that the concatenation condition would be violated, and moreover if some new box is added to a column of depth greater than $n/2$, then clearly, at the end of the process, less than E cells will be in columns of depth at most $n/2$, and moreover less than F cells will be missing from columns of depth greater than $n/2$, so the irreducible constituent of $\mathbb{S}_\beta(V) \otimes \mathbb{S}_\alpha(V)$ which is obtained, will have level less than d. □

We note the following consequence.

Lemma 6.20. *Let λ be a partition with $n-1$ rows. Let α be its efficient truncation and β its dually-efficient truncation. Write $E = \sum_{i=1}^{\lceil n/2 \rceil} i \cdot a_i$ and $F = \sum_{i=1}^{\lceil n/2 \rceil - 1} i a_{n-i}$. Then the representation $\mathbb{S}_\alpha(V)$ is a subrepresentation of $V^{\otimes E}$, the representation $\mathbb{S}_\beta(V)$ is a subrepresentation of $(V^*)^{\otimes F}$, and the representation $\mathbb{S}_\lambda(V)$ is a subrepresentation of*

$$V^{\otimes E} \otimes (V^*)^{\otimes F}.$$

Proof. The first assertion, concerning $\mathbb{S}_\alpha(V)$, is immediate from the construction of the Weyl module, since α is a partition of the integer E. The second assertion, concerning $\mathbb{S}_\beta(V)$, follows from the first, together with the fact that taking duals reverses the step vector. The third assertion, concerning $\mathbb{S}_\lambda(V)$, now follows from the previous lemma. □

Lemma 6.21. *For all $0 \leqslant d < n/2$, we have $\mathcal{L}_d = \tilde{\mathcal{L}}_d$.*

Proof. By Lemma 6.20, we have $\tilde{\mathcal{L}}_d \subseteq \mathcal{L}_d$ for all $0 \leqslant d < n/2$. It therefore suffices to show that for each $0 \leqslant d < n/2$,

$$\bigcup_{i=0}^{d} \mathcal{L}_i \subseteq \bigcup_{i=0}^{d} \tilde{\mathcal{L}}_i.$$

So suppose that ρ is an irreducible representation of $\mathrm{SU}(n)$ of level at most d. Our goal is to show that the total level of ρ is at most d. Let $E+F \leqslant d$ be such that ρ is a subrepresentation of $V^{\otimes E} \otimes (V^*)^{\otimes F}$. Just as in the proof of Lemma 6.1, we may decompose $V^{\otimes E}$ into a direct sum of submodules of the form $\mathbb{S}_\alpha(V)$ with $\alpha \vdash E$, and likewise we can decompose $(V^*)^{\otimes F}$ into a direct sum of submodules of the form $\mathbb{S}_\beta(V^*)$ with $\beta \vdash F$. Hence, we may assume that ρ is isomorphic to a subrepresentation of $\mathbb{S}_\alpha(V) \otimes \mathbb{S}_\beta(V^*) \cong \mathbb{S}_\alpha(V) \otimes (\mathbb{S}_\beta(V))^*$ for some $\alpha \vdash E$ and $\beta \vdash F$. Now as $d < n/2$, both $\mathbb{S}_\alpha(V)$ and $\mathbb{S}_\beta(V)$ are efficient, so $(\mathbb{S}_\beta(V))^*$ is dually efficient, and therefore we may apply Lemma 6.19 to deduce that the total level of ρ is at most the sum of the total levels of $\mathbb{S}_\alpha(V)$ and $(\mathbb{S}_\beta(V))^*$, which is $E+F$, as required. $\square$

Comfortable d-juntas

We say that a monomial in the matrix entries of $X \in \mathrm{SU}(n)$ is a *comfortable d-junta* if for some choice of permutation $\sigma \in S_d$ and $q_1, \ldots, q_d \in \{\textbf{imaginary}, \textbf{real}\}$ it is equal to the product $\prod_{i=1}^d (x_{i\sigma(i)})_{q_i\text{-part}}$. We say that a polynomial is a *comfortable d-junta* if it is a linear combination of comfortable monomials. We start by showing that the comfortable d-juntas are in $V_{=d}$, i.e. that they are degree d polynomials that are orthogonal to all polynomials of degree $\leqslant d-1$. (Here, the degree is in terms of the matrix entries and their complex conjugates, or equivalently, in terms of the real and the imaginary parts of the matrix entries.)

Lemma 6.22. *Every comfortable d-junta belongs to $V_{=d}$.*

Proof. Every comfortable d-junta clearly lies in $V_{\leqslant d}$. It suffices to show that any such is orthogonal to all polynomials (in the matrix entries and their complex conjugates) of degree $\leqslant d-1$. Let T be one of the degree d monomials that appear in our comfortable d-junta of degree $\leqslant d$. Let S be an arbitrary monomial of degree $\leqslant d-1$. It suffices to show that S and T are orthogonal. Let $k_1, \ldots, k_d$ be the rows of the matrix-entry variables appearing in the monomial T. Since S is a monomial of degree less than d, not all of these rows can appear amongst the variables in S; without loss of generality, assume row k_1 does not. Let k_0 be a row that appears neither in the variables in S nor those in T (such exists, since $d \leqslant n/2$, so $d + (d-1) < n$). Let U_0 be the diagonal unitary matrix with ones everywhere on the diagonal except in the (k_0, k_0) and (k_1, k_1) entries, with a $-i$ in the (k_0, k_0) entry and an i in the (k_1, k_1) entry. If X is distributed according to the Haar measure on $\mathrm{SU}(n)$, then so is $U_0 X$, but multiplying X by U_0 simply multiplies $\overline{S}T$ by i, so $\mathbb{E}_X[\overline{S}(X)T(X)] = \mathbb{E}_X[\overline{S}(U_0X)T(U_0X)] = i\mathbb{E}_X[\overline{S}(X)T(X)]$ and therefore $\mathbb{E}_X[\overline{S}T] = 0$ as required. $\square$

Lemma 6.23. *Let ρ be a representation of* $\mathrm{SU}(n)$ *of total level d, where* $0 \leqslant d \leqslant n$. *Then W_ρ contains a comfortable d-junta.*

Proof. Let ρ be an irreducible representation of $\mathrm{SU}(n)$ of total level d. Let λ be the corresponding integer partition (with at most $n-1$ rows). Let α (respectively β) be its efficient part and dually efficient part respectively. Note that the Young diagram of β^* has F cells, and all are in columns of depth at most $n/2$. As in the $\mathrm{SO}(n)$ case, there exists a comfortable, homogeneous polynomial P of total degree E, such that $P \in W_{\rho_\alpha}$, namely the polynomial

$$P_\alpha(X) := \langle X((e_1 \otimes \ldots \otimes e_d)c_\alpha), e_1 \otimes \ldots \otimes e_d \rangle,$$

where c_α is the Young symmetrizer corresponding to α; writing $c_\alpha = \sum_{\sigma \in S_d} \varepsilon_\sigma \sigma$, where $\varepsilon_\sigma \in \{-1, 0, 1\}$ for each $\sigma \in S_d$, we have

$$P_\lambda(X) = \sum_{\sigma \in S_d} \varepsilon_\sigma \prod_{i=1}^{d} x_{i\sigma(i)}.$$

Similarly, there exists a non-zero, comfortable, homogeneous polynomial Q of total degree F, such that $Q \in W_{\rho_{\beta^*}}$. Since $E + F = d \leqslant n$, we may further take P to depend only upon matrix entries in the top E by E minor, and Q to depend only upon matrix entries in the minor $[E+1, E+F] \times [E+1, E+F]$.

Note that $P\overline{Q}$ is a comfortable d-junta and is spanned by the matrix coefficients of $\rho_\alpha \otimes \rho_\beta$. We claim that, in fact, $P\overline{Q}$ is spanned by the matrix coefficients of ρ_λ. This follows immediately from the fact that, firstly, by Lemma 6.19, $\rho_\alpha \otimes \rho_\beta$ can be decomposed into a direct sum of a copy of ρ_λ and some other irreducible representations of total level less than d, and that secondly, by Lemma 6.22, the comfortable d-junta $P\overline{Q}$ is orthogonal to $V_{\leqslant d-1}^{\mathrm{SU}(n)}$.

□

6.7 Obtaining strong quasirandomness

Lemma 6.24. *Let $k, n \in \mathbb{N}$ with $k \leqslant n/2$, and let*

$$P = P(X_{11}, X_{12}, \ldots, X_{kk}) \in \mathbb{C}[X_{11}, X_{12}, \ldots, X_{kk}] \setminus \{0\}$$

be a multivariate polynomial in the variables $(X_{i,j})_{i,j \in [k]}$ that is not the zero polynomial. Let $\pi : \mathrm{SU}(n) \to \mathbb{C}^{k \times k}$ denote projection onto the top-left k by k minor of a matrix in $\mathrm{SU}(n)$. Then $P \circ \pi$ cannot vanish on all of $\mathrm{SU}(n)$.

Proof. The image of π is easily seen to have a nonempty interior inside $\mathbb{C}^{d \times d}$ (indeed, a suitably small open neighbourhood of 0 is contained in the image of π). Since P is a nontrivial polynomial in the variables $X_{11}, X_{12}, \ldots, X_{kk}$, it cannot vanish on all of a nonempty open subset of $\mathbb{C}^{k \times k}$. □

The following lemma is analogous to Lemma 6.6, and allows us to lower-bound the dimensions of irreducible representations of not-too-large degree. We prove it by considering comfortable d-juntas.

Lemma 6.25. *If ρ is an irreducible representation of $\mathrm{SU}(n)$ of level d where $0 \leqslant d < n/2$, then*

$$\dim(\rho) \geqslant \binom{\lfloor n/2 \rfloor}{d}.$$

Proof. Let P be a non-zero comfortable d-junta contained in W_ρ. Recall that the left action of $\mathrm{SU}(n)$ on $L^2(\mathrm{SU}(n))$ is defined as follows: for $A \in \mathrm{SU}(n)$ and $f \in L^2(\mathrm{SU}(n))$, we define $L_A f \in L^2(\mathrm{SU}(n))$ by $L_A f(X) = f(AX)$. Since $P \in W_\rho \setminus \{0\}$, the set $\{L_A P : A \in \mathrm{SU}(n)\}$ is contained in a left submodule V of $L^2(\mathrm{SU}(n))$ which is isomorphic to the representation ρ. In particular, for any even permutation σ, $L_{A(\sigma)}P \in V$, where $A(\sigma)$ is the permutation matrix corresponding to σ; explicitly, $(A_\sigma)_{i,j} = 1_{\{\sigma(i)=j\}}$ for each $i,j \in [n]$. Note that $L_{A(\sigma)}P$ is the polynomial obtained from P by replacing the variable $X_{i,j}$ with the variable $X_{\sigma(i),j}$, for all $i,j \in [d]$; write $\sigma P := L_{A(\sigma)}P$, for brevity. For each subset $S \in \binom{\lfloor n/2 \rfloor}{d}$, choose an even permutation σ_S sending $[d]$ to S. The polynomials $\sigma_S P$ are linearly independent as polynomials (as for any two distinct sets $S \neq S'$, the set of monomials appearing in $\sigma_S P$ is disjoint from the set of monomials appearing in $\sigma_{S'} P$); moreover, each polynomial $\sigma_S P$ depends only upon variables in the top left $[\lfloor n/2 \rfloor]$ by $[\lfloor n/2 \rfloor]$ minor. It follows from the previous lemma, applied with $k = \lfloor n/2 \rfloor$, that the polynomials $\sigma_S P$ are linearly independent as elements of $L^2(G)$, and therefore $\dim(V) \geqslant \binom{\lfloor n/2 \rfloor}{d}$, as required.

□

We remark that above method for lower-bounding the dimensions of the irreducible representations of $\mathrm{SU}(n)$ works just as well for $\mathrm{SO}(n)$ and $\mathrm{Sp}(n)$, so we could have used it in place of Lemmas 6.6 and 6.15 to give an alternative, self-contained proof of the c-strong-quasirandomness of $\mathrm{SO}(n)$ and of $\mathrm{Sp}(n)$.

The following lemma is analogous to Lemma 6.7 and lower-bounds the dimensions of irreducible representations of $\mathrm{SU}(n)$ with high levels. We defer the proof till the Appendix.

Lemma 6.26. *There exists an absolute constant $c > 0$ such that if $d \geqslant n/2$ and ρ is an irreducible representation of $\mathrm{SU}(n)$ of level d, then $\mathsf{dim}(\rho) \geqslant 2^{cn}$.*

Lemmas 6.25 and 6.26 immediately yield the strong quasirandomness of $\mathrm{SU}(n)$.

Lemma 6.27. *For each $n \geqslant 2$, the group $\mathrm{SU}(n)$ is c-strongly-quasirandom as an n-graded group, where $c > 0$ is an absolute constant.*

7 Simply connected compact Lie groups are fine

In this section, we prove Theorem 3.14. The proof has two parts. In the first part, we identify a natural noise operator U_δ on $L^2(G)$ for the groups $G = \mathrm{SU}(n), \mathrm{Sp}(n), \mathrm{Spin}(n)$ which is guaranteed to satisfy a certain hypercontractive inequality, thanks to the fact that the Ricci curvature of these groups is bounded from below. (We note that this noise operator U_δ is not quite the same as the Beckner operator $T_{\delta,r}$ that we defined earlier.) The second part consists of inferring the weak hypercontractivity of the operator $T_{\delta,r}$ from the hypercontractive inequality for U_δ. We accomplish that by analyzing the eigenvalues of U_δ and showing that they are all larger than the eigenvalues of the operator $T_{\delta^C,r}$ for some absolute constant $C > 1$. This will allow us to write $T_{\delta^C,r} = U_\delta S$ for a linear operator S on $L^2(G)$ satisfying $\|S\|_{2\to 2} \leqslant 1$. We will thus have

$$\|T_{\delta^C,r}\|_{2\to q} \leqslant \|S\|_{2\to 2} \cdot \|U_\delta\|_{2\to q} \leqslant 1,$$

as needed.

7.1 The hypercontractive inequality

Here we rely on concepts from differential geometry, such as a Riemannian metric, the Laplace–Beltrami operator, and the Ricci curvature/tensor. We use the notation of Anderson, Guionnet and Zeitouni [1, Sections E and F], and we refer the reader to that work for more details.

The simple compact Lie groups are equipped with a unique (up to normalization) structure of a bi-invariant Riemannian manifold (M, g). Once a normalization is set, and denoting by Δ the Laplace–Beltrami operator, it is also known that the Hilbert space $L^2(M)$ has an orthonormal basis of eigenvectors of Δ, that Δ is self-adjoint and negative semidefinite, and that 0 is a simple eigenvalue of Δ (with the constant functions as corresponding eigenvectors).

For a given compact Lie group G with a Riemann manifold structure, we let $u_0, u_1, u_2, \ldots$ be such a basis, with $0 = -\lambda_0 > -\lambda_1 \geqslant -\lambda_2 \geqslant \ldots$ being the corresponding eigenvalues, so that $\lambda_i > 0$ for all $i \geqslant 1$, and with u_0 being the constant function with value 1. For any $f \in L^2(M)$, we may write f uniquely in the form

$$f = \sum_{i=0}^{\infty} c_i u_i,$$

where $c_i \in \mathbb{R}$ for each $i \geqslant 0$ (we have $c_i = \langle f, u_i \rangle$ for each $i \geqslant 0$). For $\delta > 0$, we define the *noise operator* U_δ by

$$U_\delta : \; L^2(M) \to L^2(M); \quad U_\delta(f) = \sum_{i=0}^{\infty} c_i \delta^{\lambda_i} u_i,$$

for $f = \sum_{i=0}^{\infty} c_i u_i$. We note that $U_{e^{-t}}$ is, in fact, the heat kernel corresponding to Δ, which is the averaging operator with respect to the Brownian motion on the corresponding manifold.

For a Reimaniann manifold (M, g) and $C > 0$, we say that (M, g) has *Ricci curvature bounded from below by* C if for all points $p \in M$, the Ricci tensor $\mathrm{Ric}_p(\cdot, \cdot)$ at p satisfies

$$\mathrm{Ric}_p(X, X) \geqslant C g_p(X, X)$$

for all tangent vectors X at p.

The hypercontractive inequality we need is the following.

Theorem 7.1. *Let $C > 0$ and let (M, g) be a compact, connected Riemann manifold whose Ricci curvature is bounded from below by C, let $2 \leqslant p \leqslant q$ and let $f \in L^p(M)$. Then*

$$\|U_\delta(f)\|_q \leqslant \|f\|_p \quad \forall\ 0 \leqslant \delta \leqslant \left(\frac{p-1}{q-1}\right)^{1/C}.$$

As explained in Klartag and Regev [37] the Bakry-Emery criterion yields a log-Sobolev inequality (as given for example in [1, Corollary 4.4.25] applied with $\Phi = 0$), which implies a hypercontractive inequality by a theorem of Gross [16, Theorem 6]).

Utilizing the Riemannian structure. The compact simple Lie groups G are known to have a unique (up to normalization) bi-invariant Riemannian manifold structure. In order to set it up one needs to assign an inner-product on the tangent space at 1 of G, namely the corresponding Lie algebra of G. The structure at all other tangent spaces is determined by that using a push-forward with respect to left multiplication by an appropriate element.

The tangent space of $\mathrm{Spin}(n)$ is the Lie algebra of $\mathrm{SO}(n)$. As in [1] we equip it with the usual Euclidean norm, i.e. the norm of a matrix is the sum of the squares of its entries. This norm gives rise to a bi-invariant metric when applying push-forward maps to extend the metric to all tangent spaces. The norm on the Lie algebras of $\mathrm{SU}(n)$ and $\mathrm{Sp}(n)$ is defined similarly by taking the sum of squares of the components of each entry.

It is well-known (see [1, 4.4.30]), that the Ricci curvature of the simply connected compact Lie groups $\mathrm{Spin}(n), \mathrm{SU}(n), \mathrm{Sp}(n)$ is bounded from below by $(n-2)/4, \frac{n}{2}, n+1$ respectively.

Using Theorem 7.1, we can prove Theorem 3.14, modulo the following lemma.

Lemma 7.2. *There exists an absolute constant C, such that the following holds. Let G be either $\mathrm{SU}(n), \mathrm{Sp}(n)$ or $\mathrm{SO}(n)$. Let $d \in \mathbb{N}$ with $0 \leqslant d < n/2$ and let $\rho \in \mathcal{L}_d$ be a representation of level d. Then the corresponding eigenvalue of the Laplace–Beltrami operator for G satisfies $\lambda_\rho \leqslant Cnd$.*

The proof of the lemma uses a formula for the eigenvalue λ_ρ in terms of the dominant weight corresponding to ρ, which is well-known and appears e.g. in Berti and Procesi [7]. We defer the proof to the Appendix.

Proof of Theorem 3.14. We have already established the strong quasirandomness for any of the groups G of the form $\mathrm{Spin}(n), \mathrm{Sp}(n)$, and $\mathrm{SU}(n)$. It remains to establish their weak hypercontractivity. By Lemma 3.24 it is sufficient to prove that $\mathrm{SO}(n), \mathrm{Sp}(n)$ and $\mathrm{SU}(n)$ are weakly hypercontractive.

By Theorem 7.1 there exists an absolute constant C_1, such that setting $\delta = \left(\frac{1}{q-1}\right)^{\frac{1}{nC_1}}$, we have $\|U_\delta\|_{2\to q} \leqslant 1$. Let $\rho \in \mathcal{L}_d$ be of level d. Then by Lemma 7.2 we obtain that the eigenvalues of U_δ given by δ^{λ_ρ} are $\geqslant q^{-Cd}$ for some absolute constant C. This implies that we may write $T_{q^{-C},r} = U_\delta \circ S$, where all the eigenvalues of S are $\leqslant 1$. Hence,

$$\left\|T_{q^{-C},r}\right\|_{2\to q} \leqslant \|U_\delta\|_{2\to q} \|S\|_{2\to 2} \leqslant 1.$$

This completes the proof of the theorem. □

8 Showing that $\mathrm{Sp}(n), \mathrm{Spin}(n)$ and $\mathrm{SU}(n)$ are good

We now outline our coupling-based approach to showing that $\mathrm{SO}(n), \mathrm{Sp}(n)$ and $\mathrm{SU}(n)$ are good. (By Lemma 3.24, the goodness of $\mathrm{Spin}(n)$ will follow from that of $\mathrm{SO}(n)$.) We focus on the case of $\mathrm{SO}(n)$, and discuss the necessary adaptations for $\mathrm{SU}(n)$ and $\mathrm{Sp}(n)$ in the Appendix.

Our coupling approach for proving hypercontractivity

Our approach is based on constructing a coupling between matrices X sampled according to the Haar measure on $\mathrm{SO}(n)$ and matrices $Y \in \mathbb{R}^{n\times n}$ whose entries are independent standard Gaussians, with the intuition that the distributions of $\sqrt{n}X$ and Y are 'locally' close to one another.

We use this coupling to define a noise operator on $L^2(\mathrm{SO}(n), \mu)$: first we use the coupling to move a function to Gaussian space, then we apply the well-known Gaussian noise operator (the Ornstein–Uhlenbeck operator), and then we go back to $\mathrm{SO}(n)$ via the coupling. It turns out that a noise operator defined this way inherits the hypercontractive property from the Gaussian noise operator (this is easy to see), so we get hypercontractivity 'for free'. The real work of the proof is to show that the eigenvalues of our noise operator on $L^2(\mathrm{SO}(n), \mu)$ that correspond to functions in $V_{=d}$ are not too small. We show these eigenvalues are at least $2^{-O(d)}$, provided $d \leqslant \delta \cdot n^{1/2}$, for a sufficiently small absolute constant $\delta > 0$.

8.1 The Gaussian noise operator, a.k.a. the Ornstein–Uhlenbeck operator

In this section, we recall the definition of the Gaussian noise operator and several of its properties. For simplicity of notation, we present this theory for functions in $L^2(\mathbb{R}^n, \gamma)$, however everything applies more generally to functions in $L^2(\mathbb{R}^{n\times m}, \gamma)$. (Here and elsewhere, we abuse notation slightly and denote by γ a Gaussian distribution, where the domain is clear from context.)

Definition 8.1. *For $\rho \in [0, 1]$, we define $U_\rho \colon L^2(\mathbb{R}^n, \gamma) \to L^2(\mathbb{R}^n, \gamma)$ by*

$$U_\rho f(X) = \mathbb{E}_{Y\sim\gamma}[f(\rho X + \sqrt{1-\rho^2})Y].$$

It is a well known fact that U_ρ is hypercontractive [32]:

Theorem 8.2. *Let $f \colon \mathbb{R}^n \to \mathbb{R}$ and let $0 \leqslant \rho \leqslant \frac{1}{\sqrt{q-1}}$. Then $\|U_\rho f\|_{L^q(\gamma)} \leqslant \|f\|_{L^2(\gamma)}$.*

Below we use U_ρ to construct an operator T_ρ over $L^2(\sqrt{n}\mathrm{SO}(n))$ which is hypercontractive, and on which we have lower bounds on the eigenvalues corresponding to low-degrees, thereby showing that the group $\mathrm{SO}(n)$ is good.

8.2 Constructing the noise operator T_ρ

In this section, we design our noise operator T_ρ on $L^2(\mathrm{SO}(n))$. En route, we define auxiliary operators that act on both $L^2(\mathbb{R}^{n\times n}, \gamma^{n\times n})$ and $L^2(\sqrt{n}\mathrm{SO}(n), \mu)$.

Left and right multiplication by matrices from SO(n)

Definition 8.3. *For a matrix $U \in \mathrm{SO}(n)$, we define the operator L_U acting both on $L^2(\mathbb{R}^{n\times n}, \gamma^{n\times n})$ and $L^2(\sqrt{n}\mathrm{SO}(n), \mu)$, as follows. For a function $f \colon \mathbb{R}^{n\times n} \to \mathbb{R}$, the function $L_U f \colon \mathbb{R}^{n\times n} \to \mathbb{R}$ is defined by*

$$L_U f(X) = f(UX).$$

For a function $f\colon \sqrt{n}\mathrm{SO}(n) \to \mathbb{R}$*, we similarly define* $L_U f(X) = f(UX)$*.*

We similarly define the operator R_V corresponding to right multiplication.

Definition 8.4. *For a matrix* $V \in \mathrm{SO}(n)$*, we define the operator* R_V *acting both on* $L^2(\mathbb{R}^{n\times n}, \gamma)$ *and* $L^2(\sqrt{n}\mathrm{SO}(n), \mu)$*, as follows. For a function* $f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$*, the function* $R_V f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ *is defined by*

$$R_V f(X) = f(XV).$$

For a function $f\colon \sqrt{n}\mathrm{SO}(n) \to \mathbb{R}$*, we similarly define* $R_V f(X) = f(XV)$*.*

The Gram–Schmidt operators

Next, we define the operators $\mathrm{T}_{\mathsf{row}}, \mathrm{T}_{\mathsf{col}}$ that capture our coupling and map $L^2(\mathrm{SO}(n), \mu)$ to $L^2(\mathbb{R}^{n\times n}, \gamma)$, as well as their adjoint operators that go in the reverse direction. To do so, we use the Gram-Schmidt process.

Fix a matrix $X \in \mathbb{R}^{n\times n}$ and let $c_1, \ldots, c_n$ be its columns. Provided $\det(X) \neq 0$ (which for a Gaussian matrix happens with probability one), we may apply the Gram-Schmidt process on $(c_1, \ldots, c_n)$ to get an orthonormal set of vectors $\tilde{c}_1, \ldots, \tilde{c}_n$. Abusing notation slightly, we define the matrix $\mathsf{GS}_{\mathsf{col}}(X) \in \sqrt{n}\mathrm{SO}(n)$ as the matrix whose i^{th} column is $\sqrt{n}\tilde{c}_i$ for all $i < n$ and whose nth column is either $\sqrt{n}\tilde{c}_n$ (if $\det(X) > 0$) or $-\sqrt{n}\tilde{c}_n$ (if $\det(X) < 0$); this is of course a (column-) dilation of the (column-) Gram-Schmidt matrix corresponding to X. Since the Gram-Schmidt process preserves the sign of the determinant, this matrix $\mathsf{GS}_{\mathsf{col}}(X)$ is indeed in $\sqrt{n}\mathrm{SO}(n)$.

Similarly, letting $r_1, \ldots, r_n$ be the rows of X, we let $\tilde{r}_1, \ldots, \tilde{r}_n$ be the resulting set of vectors by applying the Gram-Schmidt process on $(r_1, \ldots, r_n)$ and define the matrix $\mathsf{GS}_{\mathsf{row}}(X)$ as the matrix whose i^{th} row is $\sqrt{n}\tilde{r}_i$ for all $i < n$, and whose nth row is either $\sqrt{n}\tilde{r}_n$ (if $\det(X) > 0$) or $-\sqrt{n}\tilde{r}_n$ (if $\det(X) < 0$).

The dilated Gram–Schmidt processes above define couplings $(X, \mathsf{GS}_{\mathsf{col}}(X))$ and $(X, \mathsf{GS}_{\mathsf{row}}(X))$ between γ and μ, and we use these to define the operators $\mathrm{T}_{\mathsf{row}}$ and $\mathrm{T}_{\mathsf{col}}$:

Definition 8.5. *We define* $\mathrm{T}_{\mathsf{row}}\colon L^2(\sqrt{n}\mathrm{SO}(n), \mu) \to L^2(\mathbb{R}^{n\times n}, \gamma)$ *as follows. For a function* $f\colon \sqrt{n}\mathrm{SO}(n) \to \mathbb{R}$*, we define* $\mathrm{T}_{\mathsf{row}} f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ *by*

$$\mathrm{T}_{\mathsf{row}} f(X) = f(\mathsf{GS}_{\mathsf{row}}(X)).$$

Definition 8.6. *We define* $\mathrm{T}_{\mathsf{col}}\colon L^2(\sqrt{n}\mathrm{SO}(n), \mu) \to L^2(\mathbb{R}^{n\times n}, \gamma)$ *as follows. For a function* $f\colon \sqrt{n}\mathrm{SO}(n) \to \mathbb{R}$*, we define* $\mathrm{T}_{\mathsf{col}} f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ *by*

$$\mathrm{T}_{\mathsf{col}} f(X) = f(\mathsf{GS}_{\mathsf{col}}(X)).$$

The operator T_ρ

The operators $\mathrm{T}_{\mathsf{col}}$ and $\mathrm{T}_{\mathsf{row}}$ allow us to move from $L^2(\mu)$ to $L^2(\gamma)$. We can also go in the reverse direction, using their adjoints. It is easy to see, using Jensen's inequality, that

$\mathrm{T}^*_{\mathsf{col}}U_\rho\mathrm{T}_{\mathsf{col}}$ has the same hypercontractive properties as U_ρ. Thus, it is natural to consider the operator $\mathrm{T}^*_{\mathsf{col}}U_\rho\mathrm{T}_{\mathsf{col}}$ as an analogue of the Gaussian noise operator, for $\sqrt{n}\mathrm{SO}(n)$. We do not know, however, how to bound from below the eigenvalues of $\mathrm{T}^*_{\mathsf{col}}U_\rho\mathrm{T}_{\mathsf{col}}$ so as to deduce Theorem 1.13. The reason is that to bound its eigenvalues, we (naturally) need some information about the eigenvectors corresponding to them, and we only know how to obtain such information from classical representation facts about $\mathrm{SO}(n)$. To use these facts (so as to ensure the eigenspaces are 'nice', and easy to analyse), it is necessary that our operator commutes with the action of $\mathrm{SO}(n)$ *from both sides*. For the operator $\mathrm{T}^*_{\mathsf{col}}U_\rho\mathrm{T}_{\mathsf{col}}$ above, one can show that it commutes with multiplication from the left, i.e. with the operators L_U, but unfortunately, it does not commute with multiplication from the right. To overcome this, we obtain commutation with the action of $\mathrm{SO}(n)$ from the right with an averaging trick, which is an analogue of the famous Weyl unitary trick.

Definition 8.7. *We set* $\mathrm{T}_\rho = \mathbb{E}_{V\sim\mathrm{SO}(n)}\left[R_V^*\mathrm{T}^*_{\mathsf{col}}U_\rho\mathrm{T}_{\mathsf{col}}R_V\right]$.

The following result asserts that T_ρ is hypercontractive, and it also gives lower bounds on its eigenvalues (which are required for deducing our level d inequalities).

Theorem 8.8. *For each $\rho\in(0,1)$, the operator T_ρ is self adjoint on $L^2(\sqrt{n}\mathrm{SO}(n),\mu)$ and has the following properties:*

1. *T_ρ commutes with both left and right multiplication by matrices from $\mathrm{SO}(n)$.*
2. *If $\rho\leqslant\frac{1}{\sqrt{q-1}}$, and $f\in L^2(\sqrt{n}\mathrm{SO}(n),\mu)$, then $\|\mathrm{T}_\rho f\|_{L^q(\mu)}\leqslant\|f\|_{L^2(\mu)}$.*
3. *There exists an absolute constant $\delta>0$, such that if $d\leqslant\delta n^{1/2}$ and $f\in V_d$, then*

$$\|\mathrm{T}_\rho f\|_{L^2(\mu)}\geqslant C^{-d}\rho^d\|f\|_{L^2(\mu)}.$$

Let use show how Theorem 8.8 immediately implies that $\mathrm{SO}(n)$ and $\mathrm{Spin}(n)$ are good.

Theorem 8.9. *There exist absolute constants $c,C>0$ such that the n-graded groups $\mathrm{SO}(n)$ and $\mathrm{Spin}(n)$ are (C,c)-good.*

Proof, given Theorem 8.8. The strong quasirandomness of $\mathrm{Spin}(n)$ and $\mathrm{SO}(n)$ was already established in Section 6. It remains to show that they are $(cn^{1/2},C)$-hypercontractive for absolute constants $C,c>0$. Let C be sufficiently large and c sufficiently small. By Lemma 3.24 it is sufficient to show that $\mathrm{SO}(n)$ is $(cn^{1/2},C)$-hypercontractive. Let $T=T_{\frac{1}{\sqrt{Cq}},cn^{1/2}}$ be the Beckner operator. It remains to show that $\|T\|_{2\to q}\leqslant 1$. Let us show that, by Theorem 8.8, we may write $T=T_{\frac{1}{\sqrt{q-1}}}S$, where $\|S\|_{2\to 2}\leqslant 1$. Indeed, as $T_{\frac{1}{\sqrt{q-1}}}$ is self adjoint and commutes with the action of $\mathrm{SO}(n)$ from both sides, it has the Peter-Weyl ideals W_ρ as its eigenspaces. By Part 3 of Theorem 8.8, the eigenvalues of $T_{\frac{1}{\sqrt{q-1}}}$ corresponding to a representations ρ of level d are larger than $\left(\frac{1}{C\sqrt{q}}\right)^d$, and thus are greater than the corresponding eigenvalue of T on W_ρ. This shows that the desired operator S exists. We may now apply Part 2 of Theorem 8.8 to obtain:

$$\|T\|_{2\to q} \leqslant \left\|T_{\frac{1}{\sqrt{q-1}}}\right\|_{2\to q} \|S\|_{2\to 2} \leqslant 1.$$

□

The operator T_ρ commutes with the action of SO(n) from both sides

In this section, we establish part 1 of Theorem 8.8. We phrase it as a lemma.

Lemma 8.10. *The operator* T_ρ *commutes with the action of* $\mathrm{SO}(n)$ *from both sides.*

Proof. We show the commuting from the left and the right separately.

Commuting from the left. For this, it suffices to show that for each $U, V \in \mathrm{SO}(n)$, the operators L_U and $R_V^* \mathrm{T}_{\mathsf{col}}^* \mathrm{T}_\rho \mathrm{T}_{\mathsf{col}} R_V$ commute. It is easy to see that L_U and R_V commute. Hence, it suffices to show that L_U and $\mathrm{T}_{\mathsf{col}}^* \mathrm{T}_\rho \mathrm{T}_{\mathsf{col}}$ commute.

First note that if $X \in \mathbb{R}^{n\times n}$ is a matrix, it holds that $\mathsf{GS}_{\mathsf{col}}(UX) = U\mathsf{GS}_{\mathsf{col}}(X)$. It follows that

$$L_U \mathrm{T}_{\mathsf{col}}^* f(X) = \mathrm{T}_{\mathsf{col}}^* f(UX) = \mathop{\mathbb{E}}_{Y:\mathsf{GS}_{\mathsf{col}}(Y)=UX}[f(Y)] = \mathop{\mathbb{E}}_{Z:\mathsf{GS}_{\mathsf{col}}(Z)=X}[f(UZ)] = \mathrm{T}_{\mathsf{col}}^* L_U f(X),$$

so L_U commutes with $\mathrm{T}_{\mathsf{col}}^*$. The adjointness immediately implies that L_U also commutes with the operator $\mathrm{T}_{\mathsf{col}}$. It is easy to see directly that L_U commutes with the operator U_ρ, and so it commutes with the composition $\mathrm{T}_{\mathsf{col}}^* \mathrm{T}_\rho \mathrm{T}_{\mathsf{col}}$.

Commuting from the right. Fix $V' \in \mathrm{SO}(n)$, then

$$R_{V'} \mathrm{T}_\rho = \mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}[R_{V'} R_V^* \mathrm{T}_{\mathsf{col}}^* \mathrm{T}_\rho \mathrm{T}_{\mathsf{col}} R_V] = \mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}[R_{VV'^t}^* \mathrm{T}_{\mathsf{col}}^* \mathrm{T}_\rho \mathrm{T}_{\mathsf{col}} R_V]$$

Making the change of variables $V \leftarrow VV'^t$, we get that

$$R_{V'} \mathrm{T}_\rho = \mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}[R_V^* \mathrm{T}_{\mathsf{col}}^* \mathrm{T}_\rho \mathrm{T}_{\mathsf{col}} R_{VV'}] = \mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}[R_V^* \mathrm{T}_{\mathsf{col}}^* \mathrm{T}_\rho \mathrm{T}_{\mathsf{col}} R_V] R_{V'} = \mathrm{T}_\rho R_{V'}.$$ □

8.3 The operator T_ρ is hypercontractive

In this section, we prove part 2 of Theorem 8.8. To do so, we first adopt a different point of view of the couplings defined by $\mathrm{T}_{\mathsf{row}}$ and $\mathrm{T}_{\mathsf{col}}$ that will often be easier for us to work with.

The 'Gaussian maker distribution'

Rather than going from $Y \sim \gamma$ to $X \sim \mu$ by applying the Gram–Schmidt process on its columns and dilating by $\sqrt{n}$ (and flipping the sign of the last column if necessary), we can go the other way and construct Y from X. This is accomplished as follows. We define a pair of independent random variables (X, G) such that XG is distributed according to γ and $X \sim \mu$. We call the distribution of G the *Gaussian maker distribution* and we abbreviate it to GMD.

Definition 8.11. *We define the* Gaussian maker distribution *to be the distribution of the upper-triangular matrix $G = (g_{ij})$ constructed as follows. First, independently choose one-dimensional Gaussians $g_{ij} \sim N(0, \frac{1}{n})$ of expectation zero and variance $1/n$, for each $i < j$. For each $i < n$, we independently choose g_{ii} to be $1/\sqrt{n}$ the (Euclidean) length of an $(n-i+1)$-dimensional Gaussian $z \sim (\mathbb{R}^{n-i+1}, \gamma)$. We also independently choose g_{nn} to be a standard Gaussian random variable, $z \sim N(0,1)$. Finally, we set $g_{ij} = 0$ for all $j < i$.*

It is clear that when we sample a matrix $Y \sim \gamma$ and apply the above (dilated) Gram-Schmidt process, we get a matrix X which is $\sqrt{n}$ times a matrix sampled from the Haar measure on $\mathrm{SO}(n)$ (provided we condition on the probability-one event that $\det(Y) \neq 0$, of course). We would like to show that $X^{-1}Y \sim$ GMD, independently of X. Indeed, this follows by choosing the columns of X and Y one after another. The first column of Y is uniformly distributed according to $(\mathbb{R}^n, \gamma)$. By rotational symmetry, its length and its normalization are independent, and therefore $XGe_1 = g_{11}Xe_1$ is indeed distributed as Ye_1. Note that Xe_2 is independent of g_{11}, as it is a uniformly random unit vector orthogonal to X_1. Thus, completing X_1 to a basis arbitrarily we obtain, by rotational invariance of the Gaussian distribution, that the correlation of Ye_2 with Xe_1 is normally distributed. After we present Ye_2 with respect to an extension of $\frac{1}{\sqrt{n}}Xe_1$ to an orthonormal basis, we see that the last $n-1$ coordinates are distributed as a random $n-1$ dimensional Gaussian. This shows that indeed Ye_2 is distributed as $g_{22}Xe_2 + g_{12}Xe_1$. Continuing in this fashion column by column (being a little careful with the last column), we see that indeed $X^{-1}Y$ is distributed according to GMD, independently of X. We thus have the following formulae for the adjoint operators $\mathrm{T}^*_{\mathsf{row}}, \mathrm{T}^*_{\mathsf{col}}$:

Lemma 8.12. *For each $f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ we have*

$$\mathrm{T}^*_{\mathsf{col}}f(X) = \mathop{\mathbb{E}}_{G\sim GMD}[f(XG)], \qquad \mathrm{T}^*_{\mathsf{row}}f(X) = \mathop{\mathbb{E}}_{G\sim GMD}\left[f(G^tX)\right].$$

In addition to Lemma 8.12 being useful on its own, it allows us to extend the definition of $\mathrm{T}^*_{\mathsf{col}}f$ to $\mathbb{R}^{n\times n}$. Indeed, abusing notations, we shall think of $\mathrm{T}^*_{\mathsf{col}}f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ by taking the same formula:

$$\mathrm{T}^*_{\mathsf{col}}f(X) = \mathop{\mathbb{E}}_{G\sim \mathrm{GMD}}[f(XG)].$$

This allows us to think of $\mathrm{T}^*_{\mathsf{col}}$ as an operator acting on Gaussian space, and we note that if $f\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ has degree at most d, then $\mathrm{T}^*_{\mathsf{col}}f$ also has degree at most d, and thus V_d is an invariant space of $\mathrm{T}^*_{\mathsf{col}}$.

Another consequence of Lemma 8.12 is that the operators $\mathrm{T}^*_{\mathsf{col}}, \mathrm{T}^*_{\mathsf{row}}$ do not increase q-norms:

Fact 8.13. *The following hold for all $q \geqslant 1$:*

1. *The operators $\mathrm{T}_{\mathsf{col}}$ and $\mathrm{T}_{\mathsf{row}}$ preserve q-norms.*
2. *The operators $\mathrm{T}^*_{\mathsf{col}}$ and $\mathrm{T}^*_{\mathsf{row}}$ cannot increase q-norms.*

$\mathrm{T}_{\mathsf{col}}$ *and* $\mathrm{T}_{\mathsf{row}}$ *preserve q-norms for each $q \geqslant 1$. Their adjoints have $(q \to q)$-norm at most one, for any $q \geqslant 1$.*

Proof. We prove both items for $\mathrm{T}_{\mathsf{col}}$, as the proof for $\mathrm{T}_{\mathsf{row}}$ is identical.

For the first item, we note that

$$\|\mathrm{T}_{\mathsf{col}} f\|^q_{L_q(\gamma)} = \mathop{\mathbb{E}}_{Y\sim\gamma}[|\mathrm{T}_{\mathsf{col}} f(Y)|^q] = \mathop{\mathbb{E}}_{Y\sim\gamma}[|f(GS_{\mathsf{col}}(Y))|^q] = \mathop{\mathbb{E}}_{X\sim\mu}[|f(X)|^q] = \|f\|^q_{L^q(\mu)}\,.$$

For the second item, we use Jensen's inequality:

$$\|\mathrm{T}^*_{\mathsf{col}} f\|^q_{L^q(\mu)} = \mathop{\mathbb{E}}_{X\sim\mu}\left[\left|\mathop{\mathbb{E}}_{G\sim \mathrm{GMD}}[f(XG)]\right|^q\right] \leqslant \mathop{\mathbb{E}}_{X,G}[|f(XG)|^q] = \mathop{\mathbb{E}}_{Y\sim\gamma}[|\mathrm{T}_{\mathsf{col}} f(Y)|^q] = \|f\|^q_{L^q(\gamma)}.$$

□

Deducing hypercontractivity

We now show that the operator T_ρ is hypercontractive, proving Part 3 of Theorem 8.8.

Lemma 8.14. *For all $0 \leqslant \rho \leqslant \frac{1}{\sqrt{q-1}}$ and $f\colon \mathrm{SO}(n) \to \mathbb{R}$ we have $\|\mathrm{T}_\rho f\|_{L^q(\mu)} \leqslant \|f\|_{L^2(\mu)}$.*

Proof. By the triangle inequality and the fact that R_V preserves the L^r norm for all r, it suffices to show that

$$\|\mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} f\|_{L^q(\mu)} \leqslant \|f\|_{L^2(\mu)}\,.$$

To see that this holds, we apply Fact 8.13 and Theorem 8.2:

$$\|\mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} f\|_{L^2(\gamma)} \leqslant \|U_\rho \mathrm{T}_{\mathsf{col}} f\|_{L^2(\gamma)} \leqslant \|\mathrm{T}_{\mathsf{col}} f\|_{L^2(\gamma)} = \|f\|_{L^2(\gamma)}\,.$$

□

9 Comfortable d-juntas on SO(n)

Recall that in Section 6 we defined the *comfortable d-juntas* on $\mathrm{SO}(n)$ to be the multilinear polynomials of the form $X \mapsto \sum_{\sigma\in S_d} a_\sigma \prod_{i=1}^d x_{i,\sigma(i)}$, for $a_\sigma \in \mathbb{R}$. We also showed that for any irreducible representation ρ of level d, where $0 \leqslant d < n/2$, the Peter-Weyl ideal W_ρ contains a comfortable d-junta. In this section we define comfortable polynomials in general (the comfortable d-juntas are a special case). We then show that some of them are eigenfunctions of $\mathrm{T}^*_{\mathsf{col}}$ (or of $\mathrm{T}^*_{\mathsf{row}}$).

We use comfortable d-juntas as they are both easy to work with, and each low degree eigenspace of T_ρ contains one; the latter is guaranteed by the following, since T_ρ commutes with the action of $\mathrm{SO}(n)$ from both sides.

Claim 9.1. *Let T be a linear operator on $L^2(\mathrm{SO}(n))$ that commutes with the action of $\mathrm{SO}(n)$ from both sides, and let $0 \leqslant d < n/2$. Then the space $V_{=d}$ is T-invariant, and each eigenspace of T inside $V_{=d}$ contains a comfortable d-junta.*

Proof. A linear map from $L^2(\mathrm{SO}(n))$ to itself that commutes with the action of $\mathrm{SO}(n)$ from both sides is precisely an $\mathrm{SO}(n) \times \mathrm{SO}(n)$-homomorphism. As each W_ρ is an irreducible $\mathrm{SO}(n) \times \mathrm{SO}(n)$-module and the W_ρ are pairwise non-isomorphic, Schur's lemma implies that any $\mathrm{SO}(n) \times \mathrm{SO}(n)$-homomorphism from $L^2(\mathrm{SO}(n))$ to itself acts as a scalar multiple of the identity when restricted to W_ρ. Since the linear operator T commutes with the action of $\mathrm{SO}(n)$ from both sides, each W_ρ is contained in an eigenspace of T. Since the Peter-Weyl ideas span $L^2(\mathrm{SO}(n))$, each eigenspace of T is a direct sum of some of the W_ρ's.

As $V_{=d}$ is a direct sum of finitely many Peter-Weyl ideals (each of which is T-invariant), $V_{=d}$ itself is T-invariant. The claim now follows from Fact 6.5, which implies that each of the Peter-Weyl constituents of $V_{=d}$ contains a comfortable d-junta. □

9.1 Comfortable polynomials

Claim 9.1 is important for us as it says that if we want to understand the eigenvalues of an operator T that commutes with the action of $\mathrm{SO}(n)$, it suffices to understand its action on low-degree polynomials. One can already carry out some non-trivial analysis of our hypercontractive operator using this observation, however to push our analysis all the way to $d = \Theta(n^{1/2})$ we need to work with a restricted class of polynomials, which we call 'comfortable polynomials'.

Definition 9.2 (Comfortable polynomial)**.** *We say a multilinear monomial in the matrix entries of $X \in \mathrm{SO}(n)$ is* comfortable *if it only contains variables from the top left $\lfloor \frac{n}{2} \rfloor \times \lfloor \frac{n}{2} \rfloor$ minor of X, and it contains at most one variable from each row and at most one variable from each column. A polynomial in the matrix entries is said to be* comfortable *if it is a linear combination of (multilinear) comfortable monomials.*

We may naturally index monomials in $L^2(\mathbb{R}^{n\times n}, \gamma)$ by multisets in $[n] \times [n]$. Given such a multiset $S = \{(i_1, j_1), \ldots, (i_r, j_r)\}$, where each pair (i_k, j_k) may appear multiple times, the corresponding monomial is $H_S(X) := \prod_{r=1}^{k} x_{i_r, j_r}$.

We now define the notion of comfortable polynomial.

Definition 9.3. *Let $S = \{(i_1, j_1), \ldots, (i_d, j_d)\}$ be a multiset in $\lfloor \frac{n}{2} \rfloor \times \lfloor \frac{n}{2} \rfloor$, and define its* transpose *by $S^t := \{(j_1, i_1), \ldots, (j_d, i_d)\}$.*

1. *We say H_S is* row comfortable *if $i_1, \ldots, i_d$ are distinct.*
2. *We say H_S is* column comfortable *if H_{S^t} is row comfortable, i.e. if $j_1, \ldots, j_d$ are distinct.*
3. *Finally, we say H_S is* comfortable *if it is both row comfortable and column comfortable.*

Definition 9.4. *We say a polynomial is* row comfortable *if it lies in the span of the row comfortable monomials; similarly, we say it is* column comfortable *if it lies in the span of the column comfortable monomials, and that it is* comfortable *if it is both row comfortable and column comfortable.*

We also need to define row- and column- comfortable d-juntas, generalizing the notion of comfortable d-juntas.

Definition 9.5. *A* row- (respectively column-) comfortable d-junta *is a row (respectively column) comfortable polynomial whose monomials contain variables only from the top left $d \times d$ minor.*

Comfortable monomials are eigenvectors

The following claim shows that the column/row comfortable monomials are eigenvectors of $\mathrm{T}^*_{\mathsf{col}}/\mathrm{T}^*_{\mathsf{row}}$.

Claim 9.6. *There exists $C > 0$ such that the following holds. Let $S = \{(i_1, j_1), \ldots, (i_d, j_d)\}$ be a multiset of elements of $[\lfloor \frac{n}{2} \rfloor] \times [\lfloor \frac{n}{2} \rfloor]$ that indexes a monomial. Then there exists $\lambda_S > 0$ such that:*

1. *If H_S is column comfortable (i.e. the j_k are distinct) then $\mathrm{T}^*_{\mathsf{col}} H_S = \lambda_S H_S$.*
2. *If H_S is row comfortable, then $\mathrm{T}^*_{\mathsf{row}} H_{S^t} = \lambda_S H_{S^t}$.*
3. *There exists an absolute constant $C > 0$ such that $\lambda_S \geqslant C^{-d}$ for all S.*
4. *If S is only supported on $[d] \times [d]$, then $|\lambda_S - 1| = O(d^2/n)$.*

Proof. It suffices to prove the first, third and fourth items, as the second follows from the first by taking transposes. By Lemma 8.12 we have $(\mathrm{T}^*_{\mathsf{col}} H_S)(X) = \mathbb{E}_{G \sim \mathrm{GMD}}[H_S(XG)]$. Using the fact that the entries of XG corresponding to different columns are independent and that each j_k appears at most once, we obtain

$$
\begin{aligned}
(\mathrm{T}^*_{\mathsf{col}} H_S)(X) &= \mathop{\mathbb{E}}_{G \sim \mathrm{GMD}}[H_S(XG)] \\
&= \mathop{\mathbb{E}}_{G \sim \mathrm{GMD}}\left[\prod_{k=1}^{d} (XG)_{i_k, j_k}\right] \\
&= \prod_{k=1}^{d} \mathop{\mathbb{E}}_{G \sim \mathrm{GMD}}[(XG)_{i_k, j_k}] \\
&= \prod_{k=1}^{d} \left(X \mathop{\mathbb{E}}_{G \sim \mathrm{GMD}}[G]\right)_{i_k, j_k}.
\end{aligned}
$$

Observe that $\mathbb{E}_{G \sim \mathrm{GMD}}[G]$ is a diagonal matrix with $(\mathbb{E}[G])_{j,j}$ being equal to $1/\sqrt{n}$ times the expectation of the length (= Euclidean norm) of an $(n - j + 1)$-dimensional standard

Gaussian random vector, for each $j < n$. For $m \in \mathbb{N}$, let $N(0, I_m)$ denote an m-dimensional standard Gaussian random vector, and let $\|N(0, I_m)\|_{\ell^2}$ denote its Euclidean norm. We have

$$(\mathrm{T}^*_{\mathsf{col}} H_S)(X) = \prod_{k=1}^{d} X_{i_k, j_k} \left(\mathop{\mathbb{E}}_{G \sim \mathrm{GMD}} [G] \right)_{j_k, j_k} = H_S(X) \prod_{k=1}^{d} \left(\mathop{\mathbb{E}}_{G \sim \mathrm{GMD}} [G] \right)_{j_k, j_k} = \lambda_S H_S(X),$$

where

$$\lambda_S = n^{-d/2} \prod_{k=1}^{d} \mathbb{E}[\|N(0, I_{n-j_k+1})\|_{\ell^2}].$$

To estimate the eigenvalues λ_S we need the following fact.

Fact 9.7. *For any $m \in \mathbb{N}$, we have $\sqrt{m} - \frac{1}{2\sqrt{m}} \leqslant \mathbb{E}\left[\|N(0, I_m)\|_{\ell^2}\right] \leqslant \sqrt{m}$.*

Proof. Let $Z_1, \ldots, Z_m \sim N(0,1)$ be independent and $Z = \sum_{i=1}^{m} Z_i^2$ so that $\|N(0, I_m)\|_{\ell^2} = \sqrt{Z}$. Then $\mathbb{E}[Z] = m$, so $\mathbb{E}[\sqrt{Z}] \leqslant \sqrt{\mathbb{E}[Z]} = \sqrt{m}$ by Cauchy-Schwarz, proving the upper bound. Secondly $\mathsf{var}(Z) = \sum_{i=1}^{m} \mathsf{var}(Z_i^2) = m$, which implies that

$$\mathbb{E}\left[(\sqrt{Z} - \sqrt{m})^2\right] = \mathbb{E}\left[\frac{(Z-m)^2}{(\sqrt{Z} + \sqrt{m})^2}\right] \leqslant \frac{1}{m} \mathsf{var}(Z) = 1.$$

Thus,

$$2m - 2\sqrt{m}\,\mathbb{E}\left[\sqrt{Z}\right] = \mathbb{E}[Z] + m - 2\sqrt{m}\,\mathbb{E}\left[\sqrt{Z}\right] \leqslant 1,$$

implying that

$$\mathbb{E}\left[\sqrt{Z}\right] \geqslant \sqrt{m} - \frac{1}{2\sqrt{m}},$$

proving the lower bound. □

Continuing the proof of the claim, the above fact yields $C^{-d} \leqslant \lambda_S \leqslant 1$ for all S. If S is supported upon $[d] \times [d]$, then it yields

$$\lambda_S \geqslant n^{-d/2} \left(\sqrt{n-d+1} - \frac{1}{2\sqrt{n-d+1}} \right)^d \geqslant 1 - O(d^2/n),$$

completing the proof of Claim 9.6. □

Projections onto comfortable subspaces

Define the operator $\Pi_{\mathsf{comf}}\colon L^2(\mathbb{R}^{n\times n},\gamma)\to L^2(\mathbb{R}^{n\times n},\gamma)$ to be orthogonal projection of f onto the linear subspace of comfortable polynomials. This projection also has a neat Fourier formula:

$$\Pi_{\mathsf{comf}} h(X) = \sum_{H_\alpha \text{ comfortable monomial}} \widehat{h}(\alpha) H_\alpha(X),$$

where $\hat{f}(\alpha) = \langle f, H_\alpha\rangle_{L^2(\gamma)}$. We also define the operator $\Pi_{\mathsf{comf},d}$, $\Pi_{\mathsf{comf,col},d}$ and $\Pi_{\mathsf{comf,row},d}$ to be the projections onto the space of comfortable, row comfortable and column comfortable polynomials of degree at most d, respectively.

9.2 Reducing part 3 of Theorem 8.8 to a statement about the low degree truncations of $\mathrm{T}_{\mathsf{col}}$

In this section, we prove Theorem 8.8 modulo the following lemma, asserting that on the space of d-comfortable juntas the operator $\mathrm{T}_{\mathsf{col}}R_V$ for a typical V preserves some of the mass of a function on its projection onto V_d:

Lemma 9.8. *There exist absolute constants $C>0$ and $\delta>0$ such that the following holds for all $d\leqslant \delta n^{1/2}$. For all comfortable d-juntas f we have*

$$\mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}\left[\left\|(\mathrm{T}_{\mathsf{col}}R_V f)^{\leqslant d}\right\|_{L^2(\gamma)}^2\right] \geqslant C^{-d}\left\|f\right\|_{L^2(\mu)}^2.$$

With Lemma 9.8 in hand, we can now complete the proof of Theorem 8.8.

Proof of Theorem 8.8 assuming Lemma 9.8. Lemmas 8.10 and 8.14 give the first two items, and in the rest of the argument we show the third item. Namely, letting $f\in V_d$ we want to show that $\|\mathrm{T}_\rho f\|_{L^2(\mu)}\geqslant (c\rho)^d\|f\|_{L^2(\mu)}$. By Claim 9.1 we may assume that f is a comfortable d-junta. By Cauchy–Schwarz, it is enough to show that $\langle \mathrm{T}_\rho f, f\rangle_{L^2(\mu)}\geqslant (c\rho)^d\|f\|_2^2$, and we next show that the last assertion follows by Lemma 9.8.

Using the self-adjointness and the fact that $U_\rho = U_{\sqrt{\rho}}^* U_{\sqrt{\rho}}$, we see that

$$\begin{aligned}\langle \mathrm{T}_\rho f, f\rangle_{L^2(\mu)} = \mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}\left[\|U_{\sqrt{\rho}}\mathrm{T}_{\mathsf{col}}R_V f\|_{L^2(\gamma)}^2\right] &\geqslant \mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}\left[\|(U_{\sqrt{\rho}}\mathrm{T}_{\mathsf{col}}R_V f)^{\leqslant d}\|_{L^2(\gamma)}^2\right]\\ &\geqslant \rho^d \mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}\left[\|(\mathrm{T}_{\mathsf{col}}R_V f)^{\leqslant d}\|_{L^2(\gamma)}^2\right]\\ &\geqslant \rho^d C^{-d}\left\|f\right\|_{L^2(\mu)}^2,\end{aligned}$$

where we used Lemma 9.8 and the fact that for each function g of degree $\leqslant d$ it holds that

$$\|U_{\sqrt{\rho}} g\|_{L^2(\gamma)} \geqslant \rho^{d/2}\|g\|_{L^2(\gamma)}.$$

□

10 Proof of Lemma 9.8

In this section we present the proof of Lemma 9.8 modulo a technical statement (Lemma 10.9). Our main ingredients are given below and show that one can approximate the L_2-norms of (row) comfortable d-juntas with respect to $(\sqrt{n}\mathrm{SO}(n), \mu)$ by those of $(\mathbb{R}^{n\times n}, \gamma)$.

10.1 Comparing $L^2(\mu)$ and $L^2(\gamma)$

The following lemmas assert that the 2-norm in $L^2(\mathrm{SO}(n))$ of a row comfortable d-junta is roughly bounded by its 2-norm in Gaussian space. We defer the proofs of the lemmas to Sections 10.3 and 10.4.

Lemma 10.1. *For all $\varepsilon > 0$ there exists $\delta > 0$ such that if $d \leqslant \delta n^{1/2}$ and f is a row comfortable d-junta, then*

$$\|f\|_{L^2(\mu)} \leqslant (1+\varepsilon)\|f\|_{L^2(\gamma)}.$$

In case the function f is comfortable, we are able to show that in fact also the other inequality holds.

Lemma 10.2. *For all $\varepsilon > 0$, there is $\delta > 0$ such that if $d \leqslant \delta n^{1/2}$ and f is a comfortable d-junta, then*

$$\|f\|_{L^2(\gamma)} \leqslant (1+\varepsilon)\|f\|_{L^2(\mu)}.$$

10.2 The main argument for Lemma 9.8

We are now ready to present the proof of Lemma 9.8.

Swapping between $\mathrm{T}_{\mathsf{row}}$ and $\mathrm{T}_{\mathsf{col}}$

The first step is to show that on the left hand side of Lemma 9.8, we can replace $\mathrm{T}_{\mathsf{col}}$ by $\mathrm{T}_{\mathsf{row}}$. The benefit of this exchange is that $\mathrm{T}_{\mathsf{row}}$ and R_V commute.

Lemma 10.3. *There exist absolute constants $C > 0, \delta > 0$ such that the following holds. Let $d < \delta n$ and let f be a comfortable d-junta. Then*

$$\mathop{\mathbb{E}}_{V\sim\mathrm{SO}(n)}\left[\|(\mathrm{T}_{\mathsf{col}}R_V f)^{\leqslant d}\|_{L^2(\gamma)}^2\right] \geqslant C^{-d} \mathop{\mathbb{E}}_{V\sim\mathrm{SO}(n)}\left[\|\Pi_{\mathsf{comf,d}}\mathrm{T}_{\mathsf{row}}R_V f\|_{L^2(\gamma)}^2\right].$$

Proof. Applying Claim 9.6, we have

$$\begin{aligned}
\|(\mathrm{T}_{\mathsf{col}} R_V f)^{\leqslant d}\|_{L^2(\gamma)}^2 &\geqslant \sum_{H_S \text{ comfortable of degree } d} \langle \mathrm{T}_{\mathsf{col}} R_V f, H_S\rangle_{L^2(\gamma)}^2 \\
&= \sum_{H_S \text{ comfortable of degree } d} \langle R_V f, \mathrm{T}^*_{\mathsf{col}} H_S\rangle_{L^2(\mu)}^2 \\
&= \lambda_S^2 \sum_{H_S \text{ comfortable of degree } d} \langle R_V f, H_S\rangle_{L^2(\mu)}^2 \\
&= \frac{\lambda_S^2}{\lambda_S'^2} \sum_{H_S \text{ comfortable of degree } d} \langle R_V f, \mathrm{T}^*_{\mathsf{row}} H_S\rangle_{L^2(\mu)}^2 \\
&\geqslant C^{-d} \sum_{H_S \text{ comfortable of degree } d} \langle \mathrm{T}_{\mathsf{row}} R_V f, H_S\rangle_{L^2(\gamma)}^2.
\end{aligned}$$

The lemma follows by plugging in the definition of $\Pi_{\mathsf{comf},d}$. □

$\mathrm{T}_{\mathsf{row}} f$ is close to f

We have thus reduced our task to understanding the average of the square of the 2-norm of $\Pi_{\mathsf{comf,d}} \mathrm{T}_{\mathsf{row}} R_V f = \Pi_{\mathsf{comf,d}} R_V \mathrm{T}_{\mathsf{row}} f$, where the transition is because $\mathrm{T}_{\mathsf{row}}$ and R_V commute. The following claim further simplifies our task and shows that $\mathrm{T}_{\mathsf{row}} f$ is close to f, thereby effectively reducing our task to estimating the 2-norm of $\Pi_{\mathsf{comf,d}} R_V f$ (some care is required to make this precise as we are applying a projection operator on top, which may decrease norms considerably).

Claim 10.4. *For each $\varepsilon > 0$, there exists $\delta > 0$, such that the following holds. Let $d \leqslant \delta n^{1/2}$, and let f be a comfortable d-junta. Then*

$$\|\mathrm{T}_{\mathsf{row}} f - f\|_{L^2(\gamma)}^2 \leqslant \varepsilon \, \|f\|_{L^2(\gamma)}^2 .$$

Proof. Let g be a comfortable d-junta satisfying $\mathrm{T}^*_{\mathsf{row}} g = f$, i.e. writing $f = \sum_S \alpha_S H_S$, we take $g = \sum \lambda_{S^t}^{-1} \alpha_S H_S$, where λ_S is as in Claim 9.6. Then by Parseval

$$\|g - f\|_{L^2(\gamma)} \leqslant \frac{\varepsilon}{10} \|f\|_{L^2(\gamma)}$$

provided that δ is sufficiently small. Hence by Cauchy–Schwarz

$$\begin{aligned}
\|f\|_{L^2(\mu)}^2 = \langle f, \mathrm{T}^*_{\mathsf{row}} g\rangle = \langle \mathrm{T}_{\mathsf{row}} f, g\rangle &= \langle \mathrm{T}_{\mathsf{row}} f, f\rangle + \langle \mathrm{T}_{\mathsf{row}} f, g - f\rangle \\
&\leqslant \langle \mathrm{T}_{\mathsf{row}} f, f\rangle + \|\mathrm{T}_{\mathsf{row}} f\|_{L^2(\gamma)} \|g - f\|_{L^2(\gamma)},
\end{aligned}$$

implying

$$\langle \mathrm{T}_{\mathsf{row}} f, f\rangle \geqslant \|f\|_{L^2(\mu)}^2 - \frac{\varepsilon}{10} \|\mathrm{T}_{\mathsf{row}} f\|_{L^2(\gamma)} \|f\|_{L^2(\gamma)} = \|f\|_{L^2(\mu)}^2 - \frac{\varepsilon}{10} \|f\|_{L^2(\mu)} \|f\|_{L^2(\gamma)}.$$

Thus, we get that

$$\begin{aligned}\|\mathrm{T}_{\mathsf{row}} f - f\|_{L^2(\gamma)}^2 &= \|f\|_{L^2(\gamma)}^2 + \|\mathrm{T}_{\mathsf{row}} f\|_{L^2(\gamma)}^2 - 2\langle \mathrm{T}_{\mathsf{row}} f, f\rangle \\ &\leqslant \|f\|_{L^2(\gamma)}^2 - \|f\|_{L^2(\mu)}^2 + \frac{\varepsilon}{5}\|f\|_{L^2(\mu)}\|f\|_{L^2(\gamma)}.\end{aligned}$$

which using Lemmas 10.1 and 10.2 is at most $\varepsilon\|f\|_{L^2(\gamma)}^2$.

□

The projection of $R_V \mathrm{T}_{\mathsf{row}} f$ onto the subspace of comfortable polynomials

With the steps we have collected so far, it seems that to finish the proof of Lemma 9.8 it suffices to estimate the typical 2-norm squared of $\Pi_{\mathsf{comf},d} R_V f$. While this is indeed the case, some care is needed as one cannot really smoothly switch $\mathrm{T}_{\mathsf{row}} f$ to f in the previous statement, and to address that we must be able to estimate the 2-norm of $\Pi_{\mathsf{comf},d} R_V f$ under the weaker hypothesis that f is a row comfortable d-junta. This is the content of the following lemma:

Claim 10.5. *Let $d \leqslant n/2$ and let f be a row-comfortable d-junta. Then*

$$\mathbb{E}_{V\sim \mathrm{SO}(n)}\|\Pi_{\mathsf{comf},d} R_V f\|_{L^2(\gamma)}^2 = \frac{(n/2)!}{n^d((n/2)-d)!}\|f\|_{L^2(\mu)}^2.$$

Proof. We first argue that

$$\mathbb{E}_{V\sim \mathrm{SO}(n)}\|\Pi_{\mathsf{comf},d} R_V f\|_{L^2(\gamma)}^2 = \frac{(n/2)!}{((n/2)-d)!}\mathbb{E}_V\langle R_V f, H_{((1,1),\dots,(d,d))}\rangle^2.$$

Indeed, the assertion follows from the fact that we may write V as the product of a random $\mathrm{SO}(n)$ matrix and a random permutation matrix. Now the left S_n orbit of a monomial of the form $\prod_{i=1}^d x_{iv_i}$ with v_i distinct consists of all such monomials.

Write $f = \sum_{\alpha=((1,i_1),\dots,(d,i_d))} \hat{f}(\alpha) H_\alpha$. Then for each such $\alpha = ((1,i_1),\dots,(d,i_d))$ we have $R_V H_\alpha(X) = H_\alpha(XV) = \prod_{k=1}^d [XV]_{k,i_k} = \sum_{r_1,\dots,r_d}\prod_{k=1}^d X_{k,r_k}V_{r_k,i_k}$. Therefore by Plancherel,

$$\langle R_V H_\alpha(X), H_{(1,1),\dots,(d,d)}\rangle = H_\alpha(V) = n^{-d/2} H_\alpha(\sqrt{n}V).$$

The equality part of the lemma follows by expanding $f = \sum_\alpha \hat{f}(\alpha) H_\alpha$ and taking 2-norms over $\sqrt{n}V \sim \mu$.

□

We now show how to lower bound the left hand side of Lemma 10.3.

Lemma 10.6. *There exist $\delta > 0$ and $C > 0$, such that the following holds. Let $d \leqslant \delta n^{1/2}$. Then for all comfortable d-juntas f we have*

$$\mathop{\mathbb{E}}_{V\sim \mathrm{SO}(n)}\left[\|\Pi_{\mathsf{comf},d} R_V \mathrm{T}_{\mathsf{row}} f\|_2^2\right] \geqslant C^{-d}\|f\|_{L^2(\mu)}^2.$$

Proof. We write $\Pi_{\mathsf{comf},d} = \Pi_{\mathsf{comf},d}\Pi_{\mathsf{comf},\mathsf{row},d}$, and set $g = \Pi_{\mathsf{comf},\mathsf{row},d}\mathrm{T}_{\mathsf{row}}f$. We have

$$\Pi_{\mathsf{comf},d}R_V\mathrm{T}_{\mathsf{row}}f = \Pi_{\mathsf{comf},d}R_V g,$$

where we used the fact that $\Pi_{\mathsf{comf},\mathsf{row},d}$ commutes with R_V. Taking the squares of the 2-norms and expectations over V we may apply Claim 10.5 (using the fact that g is a row comfortable d-junta) to obtain that

$$\mathop{\mathbb{E}}_{V}\left[\|\Pi_{\mathsf{comf},d}\mathrm{T}_{\mathsf{row}}R_V f\|_{L^2(\gamma)}^2\right] = \frac{(n/2)!}{n^d((n/2)-d)!}\|g\|_{L^2(\mu)}^2 \geqslant C^{-d}\|g\|_{L^2(\mu)}^2. \tag{11}$$

By Lemma 10.1, as $g-f$ is row comfortable we have

$$\|g\|_{L^2(\mu)} \geqslant \|f\|_{L^2(\mu)} - \|g-f\|_{L^2(\mu)} \geqslant \|f\|_{L^2(\mu)} - 2\|g-f\|_{L^2(\gamma)}$$

As f is comfortable we can apply Claim 10.4 to obtain

$$\|g-f\|_{L^2(\gamma)} = \|\Pi_{\mathsf{comf},\mathsf{row},d}(\mathrm{T}_{\mathsf{row}}f - f)\|_{L^2(\gamma)} \leqslant \|\mathrm{T}_{\mathsf{row}}f - f\|_{L^2(\gamma)} \leqslant \frac{\varepsilon}{10}\|f\|_{L^2(\gamma)}.$$

On the other hand Lemma 10.2 shows that $\|f\|_{L^2(\gamma)} \leqslant 2\|f\|_{L^2(\mu)}$. Putting everything together we obtain that $\|g\|_{L^2(\mu)} > (1-\varepsilon/2)\|f\|_{L^2(\mu)}$. Plugging this into (11) completes the proof. □

Finishing the proof of Lemma 9.8

Using Lemma 10.3, the left hand side of Lemma 9.8 is at least

$$C^{-d}\mathop{\mathbb{E}}_{V\sim\mathrm{SO}(n)}\left[\|\Pi_{\mathsf{comf,d}}\mathrm{T}_{\mathsf{row}}R_V f\|_{L^2(\gamma)}^2\right] = C^{-d}\mathop{\mathbb{E}}_{V\sim\mathrm{SO}(n)}\left[\|\Pi_{\mathsf{comf,d}}R_V\mathrm{T}_{\mathsf{row}}f\|_{L^2(\gamma)}^2\right]$$

and using Lemma 10.6 the last quantity is at least $C'^{-d}\,\|f\|_{L^2(\mu)}^2$ as required. □

10.3 $L^2(\mu)$ is dominated by $L^2(\gamma)$: Proof of Lemma 10.1

In the following two sections, we prove Lemmas 10.1 and 10.2, modulo Lemma 10.9 which is proved in Section 11. First, we show that the operator $\mathrm{T}^*_{\mathsf{col}}$ is close to the identity on column comfortable d-juntas.

Claim 10.7. *For all $\varepsilon > 0$ there exists $\delta > 0$ such that if $d \leqslant \delta n^{1/2}$ and f is a column comfortable d-junta, then*

$$\|\mathrm{T}^*_{\mathsf{col}}f - f\|_{L^2(\gamma)} \leqslant \varepsilon\|f\|_{L^2(\gamma)}.$$

Proof. The lemma follows immediately from Parseval and Claim 9.6 as $|\lambda_S - 1| < \varepsilon$ for each S, provided that δ is sufficiently small (similarly to in the proof of Claim 10.4). □

Claim 10.7 is particularly useful as it implies that $\mathrm{T}^*_{\mathsf{col}}$ is invertible on the space of column comfortable d-juntas, and thus gives us a natural way of going from $L^2(\mu)$ to $L^2(\gamma)$. We are now ready to prove Lemma 10.1, restated below.

Lemma 10.1 (Restated) . *For all $\varepsilon > 0$, there exists $\delta > 0$ such that if $d \leqslant \delta n^{1/2}$, then for any function f that is either a column comfortable d-junta or a row comfortable d-junta, we have*

$$\|f\|_{L^2(\mu)} \leqslant (1+\varepsilon)\|f\|_{L^2(\gamma)}.$$

Proof. Without loss of generality, we assume that f is column comfortable; otherwise we may consider $f'(X) = f(X^t)$. Let g be a function such that $\mathrm{T}^*_{\mathsf{col}} g = f$. Then as $\mathrm{T}^*_{\mathsf{col}}$ is a contraction, we have $\|f\|_{L^2(\mu)} \leqslant \|g\|_{L^2(\gamma)}$. In order to complete the proof, we note that by Claim 10.7 the operator $\mathrm{T}^*_{\mathsf{col}}$ is invertible on column comfortable d-junta and its inverse has 2-norm at most $1+\varepsilon$. We have $g = (\mathrm{T}^*_{\mathsf{col}})^{-1} f$, and so $\|g\|_{L^2(\gamma)} = \left\|(\mathrm{T}^*_{\mathsf{col}})^{-1} f\right\|_{L^2(\gamma)} \leqslant (1+\varepsilon)\left\|f\right\|_{L^2(\gamma)}$. Indeed, otherwise we may find a column comfortable d-junta h of 2-norm 1 such that $\left\|{\mathrm{T}^*_{\mathsf{col}}}^{-1} h\right\|_2 > 1+\varepsilon$, and then for $h' = (\mathrm{T}^*_{\mathsf{col}})^{-1} h$ (which is a column comfortable d-junta by Claim 9.6) we get:

$$\left\|\mathrm{T}^*_{\mathsf{col}} h' - h'\right\|_{L^2(\gamma)} = \left\|h - {\mathrm{T}^*_{\mathsf{col}}}^{-1} h\right\|_{L^2(\gamma)} \geqslant \left\|{\mathrm{T}^*_{\mathsf{col}}}^{-1} h\right\|_{L^2(\gamma)} - \|h\|_{L^2(\gamma)} > 1+\varepsilon-1 = \varepsilon,$$

and contradiction to Claim 10.7. □

10.4 $L^2(\gamma)$ is dominated by $L^2(\mu)$: Proof of Lemma 10.2

To prove Lemma 10.2 we define an auxiliary distribution over $\mathbb{R}^{n\times n}$ which we refer to as the 'over-Gaussian' distribution:

Definition 10.8. *Let $G \sim GMD$, and choose $Y \sim \gamma$ independently. We define the distribution ν to be the distribution of YG, and call it the* over-Gaussian distribution*.*

We refer to ν by this name since it can be thought of as taking an $X \sim \mu$ an $\mathrm{SO}(n)$-matrix and then multiplying it by two independent copies of GMD, thereby "overshooting" the Gaussian distribution.

In the following section, we show that the distribution ν is close to γ in the sense that the expectation of certain test functions remain roughly the same under both measures. In fact, the test functions of interest for us are the squares of the comfortable d juntas and they do remain roughly the same even though we allow the degree to be up to $\Omega(\sqrt{n})$. More precisely, the following lemma asserts that if f is a comfortable d-junta, then its over-Gaussian 2-norm cannot be much larger than its Gaussian 2-norm.

Lemma 10.9. *For all $\varepsilon > 0$ there exists $\delta > 0$ such that if $d \leqslant \delta n^{1/2}$, then for all comfortable d-juntas f, we have*

$$\|f\|_{L^2(\nu)} \leqslant (1+\varepsilon)\|f\|_{L^2(\gamma)}.$$

With Lemma 10.9 in hand, the proof of Lemma 10.2 (restated below) readily follows.

Lemma 10.2 (Restated) . *For all $\varepsilon > 0$, there exists $\delta > 0$ such that if $d \leqslant \delta n^{1/2}$ and f is a comfortable d-junta, then $\|f\|_{L^2(\gamma)} \leqslant (1+\varepsilon)\|f\|_{L^2(\mu)}$.*

Proof. We may assume that $\varepsilon \leqslant 1/2$. By the triangle inequality it is sufficient to show that $\|f(X) - f(XG)\|_{L^2(\mu,\mathrm{GMD})} < \frac{\varepsilon}{2}\|f\|_{L^2(\gamma)}$, where $X \sim \mu$ and $G \sim \mathrm{GMD}$. Since for any fixed, upper triangular $G \in \mathbb{R}^{n\times n}$, the function g_G defined by

$$g_G : X \mapsto f(X) - f(XG)$$

is a row-comfortable d-junta, by Lemma 10.1 we have

$$\begin{aligned}
\|f(X) - f(XG)\|^2_{L^2(\mu,\mathrm{GMD})} &= \mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{X\sim\mu}[(f(X) - f(XG))^2] \\
&= \mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{X\sim\mu}[g_G(X)^2] \\
&= \mathbb{E}_{G\sim\mathrm{GMD}}[\|g_G\|^2_{L^2(\mu)}] \\
&\leqslant (1+\varepsilon)^2\,\mathbb{E}_{G\sim\mathrm{GMD}}\|g_G\|^2_{L^2(\gamma)} \\
&= (1+\varepsilon)^2\,\mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[(f(Y) - f(YG))^2] \\
&= (1+\varepsilon)^2\,\|f(Y) - f(YG)\|^2_{L^2(\gamma,\mathrm{GMD})},
\end{aligned}$$

and therefore it suffices to show that

$$\|f(Y) - f(YG)\|_{L^2(\gamma,\mathrm{GMD})} < \frac{\varepsilon}{4}\|f\|_{L^2(\gamma)},$$

where $Y \sim \gamma$ and $G \sim \mathrm{GMD}$ is independent of Y. Expanding, we note that

$$\|f(Y) - f(YG)\|^2_{L^2(\gamma,\mathrm{GMD})} = \|f\|^2_{L^2(\gamma)} + \|f\|^2_{L^2(\nu)} - 2\langle f, \mathrm{T}^*_{\mathsf{col}} f\rangle_{L^2(\gamma)}.$$

To handle the cross term, we note that by Cauchy-Schwarz and Claim 10.7, we have

$$\left|\langle f, \mathrm{T}^*_{\mathsf{col}} f - f\rangle_{L^2(\gamma)}\right| \leqslant \|f\|_{L^2(\gamma)}\,\|\mathrm{T}^*_{\mathsf{col}} f - f\|_{L^2(\gamma)} \leqslant \frac{\varepsilon}{10}\,\|f\|^2_{L^2(\gamma)},$$

so $\langle f, \mathrm{T}^*_{\mathsf{col}} f\rangle_{L^2(\gamma)} \geqslant (1-\varepsilon/10)\|f\|^2_{L^2(\gamma)}$. Using Lemma 10.9, we have

$$\|f\|^2_{L^2(\nu)} \leqslant (1+\varepsilon/20)\|f\|^2_{L^2(\gamma)}$$

provided δ is sufficiently small, and so

$$\|f(Y) - f(YG)\|^2_{L^2(\gamma,\mathrm{GMD})} \leqslant \|f\|^2_{L^2(\gamma)} + \|f\|^2_{L^2(\nu)} - 2(1-\varepsilon/10)\|f\|^2_{L^2(\gamma)} \leqslant (\varepsilon/4)\|f\|^2_{L^2(\gamma)},$$

completing the proof. □

11 Proof of Lemma 10.9

Our aim in this section is to prove Lemma 10.9. We begin with some more notation. For a permutation $\sigma \in S_d$, we write $x_\sigma := x_{1,\sigma(1)} \cdots x_{d,\sigma(d)}$; this is a function on $\mathbb{R}^{d\times d}$. Fix a comfortable d-junta and write $f = \sum_{I=(i_1,\ldots,i_d)} a_I x_I$ where $x_I := x_{1,i_1}x_{2,i_2}\cdots x_{d,i_d}$, the sum ranging over all I such that $i_1,\ldots,i_d \in [d]$ are distinct. (As usual, we write $(S)_d$ for the set of ordered d-tuples of distinct elements of the set S, so we write $I \in ([d])_d$.) We prove the following two claims, which handle respectively the diagonal terms and the off-diagonal terms when computing $\|f\|_{L^2(\nu)}$. For the diagonal terms we have:

Claim 11.1. *If $i_1, \dots, i_d \in [n]$ are distinct, then $x_I := x_{1,i_1} x_{2,i_2} \cdots x_{d,i_d}$ satisfies $\|x_I\|^2_{L^2(\nu)} = 1$.*

The second claim deals with the off-diagonal terms. For $I, J \in [n]^d$ we let $d(I, J) := |\{ r \mid i_r \neq j_r \}|$ denote the Hamming distance from I to J. We have

Claim 11.2. *For any I, J such that $d(I, J) = \ell$, we have $|\langle x_I, x_J \rangle|_{L^2(\nu)} \leqslant \varepsilon_\ell$, where*

$$\varepsilon_\ell := 2^{\ell+4} n^{-\ell/2} 2^{d\ell/\sqrt{n}}.$$

11.1 Claims 11.1 and 11.2 imply Lemma 10.9

We first show how to deduce Lemma 10.9 from Claims 11.1 and 11.2. Expanding, we get that

$$\begin{aligned}
\|f\|^2_{L^2(\nu)} &= \sum_I \alpha_I^2 \|x_I\|^2_{L^2(\nu)} + \sum_{I \neq J} a_I a_J \langle x_I, x_J \rangle_\nu \\
&\leqslant \|f\|^2_{L^2(\gamma)} + \sum_{I \neq J} \frac{a_I^2 + a_J^2}{2} |\langle x_I, x_J \rangle_\nu| \\
&\leqslant \|f\|^2_{L^2(\gamma)} + \sum_{\ell=1}^{d} \sum_I a_I^2 \, |\{J : d\,(J, I) = \ell\}| \cdot \varepsilon_\ell \\
&\leqslant \|f\|^2_{L^2(\gamma)} + \|f\|^2_{L^2(\gamma)} \sum_{\ell=1}^{d} \varepsilon_\ell \binom{d}{\ell} \ell! \\
&= \|f\|^2_{L^2(\gamma)} \left(1 + \sum_{\ell=1}^{d} \varepsilon_\ell d^\ell \right).
\end{aligned}$$

Using the upper bound on ε_ℓ we get that

$$\sum_{\ell=1}^{d} \varepsilon_\ell d^\ell \leqslant \sum_{\ell=1}^{d} 2^{\ell+4} n^{-\ell/2} 2^{d\ell/\sqrt{n}} d^\ell \leqslant 16 \sum_{\ell=1}^{\infty} \left(\frac{2d \cdot 2^{d/\sqrt{n}}}{\sqrt{n}} \right)^\ell \leqslant \frac{\varepsilon}{2},$$

where we used $d \leqslant \delta n^{1/2}$ and the fact that δ is sufficiently small compared to ε.

11.2 Proof of Claims 11.1, 11.2

To prove the two claims, we first need the following simple computation regarding the Gaussian maker distribution:

Claim 11.3. *Let $I = (i_1, \dots, i_d) \in [n]^d$ and $J = (j_1, \dots, j_d) \in [n]^d$ be such that $d(I, J) = \ell$, and such that in the product*

$$G_{i_1 j_1} G_{i_2 j_2} \cdots G_{i_d j_d},$$

no matrix entry of G appears more than twice. Then

$$|\mathbb{E}_{G\sim GMD}\left[G_{i_1j_1}G_{i_2j_2}\cdots G_{j_di_d}\right]| \leqslant \left(\frac{1}{n}\right)^{\ell/2}.$$

Proof. If, in the product

$$G_{i_1j_1}G_{i_2j_2}\cdots G_{i_dj_d},$$

some off-diagonal matrix entry of G appears exactly once, then the expectation of the product is zero. We may therefore assume that every off-diagonal matrix entry of G appears either exactly twice, or not at all, in the above product. If there are exactly ℓ values of r such that $i_r \neq j_r$, then the above expectation factorises into a product of the expectations of the squares of $\ell/2$ off-diagonal and of the squares of $(d-\ell)/2$ diagonal entries:

$$\prod_{k\in\mathcal{D}}\mathbb{E}[G_{k,k}^2]\prod_{(i,j)\in\mathcal{E}}\mathbb{E}[G_{i,j}^2],$$

where $\mathcal{E}\subset[n]^2\setminus\{(k,k):\ k\in[n]\}$, $|\mathcal{D}|=(d-\ell)/2$ and $|\mathcal{E}|=\ell/2$. We have $\mathbb{E}[G_{i,j}^2]=1/n$ for all $(i,j)\in\mathcal{E}$ and $\mathbb{E}[G_{k,k}^2]=(n-k+1)/n\leqslant 1$ for all $k\in\mathcal{D}$, proving the claim. □

We are now ready to prove Claim 11.1.

Proof of Claim 11.1. Let $x_I = x_{1,i_1}x_{2,i_2}\cdots x_{d,i_d}$, where $i_1,\ldots,i_d\in[n]$ are distinct. We have

$$\|x_I\|_{L^2(\nu)}^2 = \mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[(YG)_{1,i_1}^2(YG)_{2,i_2}^2\cdots(YG)_{d,i_d}^2].$$

Since for each $h\in[d]$, $(YG)_{h,i_h}=\sum_{k=1}^{i_h}Y_{h,k}G_{k,i_h}$ involves only entries of Y in row h and entries of G in column i_h (and the i_h are distinct), the random variables $((YG)_{h,i_h}^2:\ h\in[d])$ form a system of independent random variables, and therefore

$$\|x_I\|_{L^2(\nu)}^2 = \mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[(YG)_{1,i_1}^2]\mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[(YG)_{2,i_2}^2]\cdots\mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[(YG)_{d,i_d}^2].$$

For each $h \in [d]$, we have

$$\begin{aligned}
\mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[(YG)^2_{h,i_h}] &= \mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}\left[\left(\sum_{k=1}^{i_h} Y_{h,k}G_{k,i_h}\right)^2\right] \\
&= 2\sum_{1\leqslant k<k'\leqslant i_h} \mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[Y_{h,k}Y_{h,k'}G_{k,i_h}G_{k',i_h}] \\
&+ \sum_{k=1}^{i_h} \mathbb{E}_{G\sim\mathrm{GMD}}\mathbb{E}_{Y\sim\gamma}[Y^2_{h,k}G^2_{k,i_h}] \\
&= 0 + \sum_{k=1}^{i_h} \mathbb{E}_{G\sim\mathrm{GMD}}[G^2_{k,i_h}]\mathbb{E}_{Y\sim\gamma}[Y^2_{h,k}] \\
&= \sum_{k=1}^{i_h} \mathbb{E}_{G\sim\mathrm{GMD}}[G^2_{k,i_h}]\cdot 1 \\
&= (i_h-1)(1/n) + (n-i_h+1)/n \\
&= 1.
\end{aligned}$$

(Here, for the third equality we use the independence of $Y_{h,k}, Y_{h,k'}, G_{k,i_h}, G_{k',i_h}$ and the fact that $Y_{h,k}$ and $Y_{h,k'}$ both have zero expectation.) Hence, $\|x_I\|^2_{L^2(\nu)} = 1$, as required. □

We now move on to the proof of Claim 11.2.

Proof of Claim 11.2. Let $\ell \geqslant 1$ and fix $I, J \in ([d])_d$ such that $d(I,J) = \ell \geqslant 1$. Since G is upper-triangular and $i_h, j_h \leqslant d$ for all $h \in [d]$, we have

$$(YG)_{h,i_h} = \sum_{k=1}^{i_h} Y_{h,k}G_{k,i_h} = \sum_{k=1}^{d} Y_{h,k}G_{k,i_h}$$

and

$$(YG)_{h,j_h} = \sum_{k=1}^{j_h} Y_{h,k}G_{k,j_h} = \sum_{k=1}^{d} Y_{h,k}G_{k,j_h}$$

for all $h \in [d]$. Hence,

$$x_I\,(YG) = \prod_{h=1}^{d}(YG)_{h,i_h} = \sum_{K=(k_1,\ldots,k_d)\in[d]^d} Y_{1,k_1}\cdots Y_{d,k_d}G_{k_1,i_1}\cdots G_{k_d,i_d}$$

and

$$x_J\,(YG) = \sum_{K=(k_1,\ldots,k_d)\in[d]^d} Y_{1,k_1}\cdots Y_{d,k_d}G_{k_1,j_1}\cdots G_{k_d,j_d},$$

so, using the fact that, under ν, the $(Y_{i,j} : i,j \in [n])$ are independent and of expectation zero (and are independent of the $G_{i,j}$), we obtain

$$\langle x_I, x_J \rangle_\nu = \sum_{K\in[d]^d} \mathbb{E}_{G\sim \mathrm{GMD}}\left[G_{k_1 i_1} G_{k_1 j_1} \cdots G_{k_d i_d} G_{k_d j_d}\right].$$

For a d-tuple $K = (k_1, \ldots, k_d) \in [d]^d$ we write $m_1 = m_1(K) := |\{r : j_r = i_r, k_r \neq i_r\}|$, $m_2 = m_2(K) := |\{r : j_r \neq i_r, k_r \notin \{i_r, j_r\}\}|$ and $m_3 = m_3(K) := |\{r : j_r \neq i_r, k_r \in \{i_r, j_r\}\}|$, and we let $\mathcal{K}(m_1, m_2, m_3)$ denote the set of d-tuples K with parameters m_1, m_2 and m_3. For $K \in \mathcal{K}(K_1, K_2, K_3)$, by Claim 11.3 we have

$$\left|\mathbb{E}_{G\sim \mathrm{GMD}}\left[G_{k_1 i_1} G_{k_1 j_1} \cdots G_{k_d i_d} G_{k_d j_d}\right]\right| \leqslant n^{-\frac{2m_1+2m_2+m_3}{2}}.$$

Summing over all K, we see that $|\langle x_I, x_J\rangle|$ is at most

$$\begin{aligned}
&\sum_{m_1,m_2,m_3} \sum_{K\in\mathcal{K}(m_1,m_2,m_3)} n^{-\frac{2m_1+2m_2+m_3}{2}} \\
&\leqslant \sum_{m_1,m_2,m_3} n^{-\frac{2m_1+2m_2+m_3}{2}} \left|\mathcal{K}(m_1,m_2,m_3)\right|.
\end{aligned}$$

Now

$$|\mathcal{K}(m_1,m_2,m_3)| \leqslant \binom{d}{m_1} d^{m_1} \binom{\ell}{m_2} d^{m_2} 2^{m_3} \leqslant \frac{d^{2m_1+m_2}\ell^{m_2}2^{m_3}}{m_1! m_2!}.$$

Summing over all m_1, m_2, m_3 with $m_2 + m_3 = \ell$ completes the proof. Indeed,

$$\begin{aligned}
\sum_{m_1,m_2,m_3} n^{-\frac{2m_1+2m_2+m_3}{2}} |\mathcal{K}(m_1,m_2,m_3)| &\leqslant \sum_{m_1,m_2,m_3} n^{-\frac{2m_1+2m_2+m_3}{2}} \frac{d^{2m_1+m_2}\ell^{m_2}2^{m_3}}{m_1! m_2!} \\
&= \sum_{m_2+m_3=\ell} n^{-\frac{2m_2+m_3}{2}} \frac{d^{m_2}\ell^{m_2}2^{m_3}}{m_2!} \sum_{m_1=0}^{d-\ell} \frac{1}{m_1!}(d^2/n)^{m_1} \\
&\leqslant 2 \sum_{m_2+m_3=\ell} n^{-\frac{2m_2+m_3}{2}} \frac{d^{m_2}\ell^{m_2}2^{m_3}}{m_2!} \\
&= 2^{\ell+1} n^{-\ell/2} \sum_{m_2=0}^{\ell} n^{-m_2/2} \frac{d^{m_2}\ell^{m_2}}{2^{m_2} m_2!} \\
&\leqslant 2^{\ell+1} n^{-\ell/2} \left(1 + \sum_{m_2=1}^{\ell} \left(\frac{ed\ell}{2\sqrt{n} m_2}\right)^{m_2}\right) \\
&\leqslant 2^{\ell+4} n^{-\ell/2} 2^{d\ell/\sqrt{n}},
\end{aligned}$$

as required. □

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices.* Cambridge University Press, Cambridge, 2010.

[2] S. Aruanachalam, U. Girish, and N. Lifshitz. One clean qubit suffices for quantum communication advantage. In preparation.

[3] L. Babai and V.T. Sós. Sidon sets in groups and induced subgraphs of Cayley graphs. *Europ. J. Combin.*, 6:101–114, 1985.

[4] D. Bakry and M. Émery. Diffusions hypercontractives. In J. Azéma and M. Yor, editors, *Séminaire de Probabilités XIX 1983/84*, pages 177–206. Springer, Berlin, Heidelberg, 1985.

[5] W. Beckner. Inequalities in Fourier Analysis. *Ann. Math.*, 102:159–182, 1975.

[6] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Publ. Math. Inst. Hautes Etudes Sci.*, 90(1):5–43, 1999.

[7] M. Berti and C. Procesi. Nonlinear wave and Schrödinger equations on compact Lie groups and homogeneous spaces. *Duke Math. J.*, 159:479–538, 2011.

[8] E. Breuillard, B. Green, and T. Tao. Approximate subgroups of linear groups. *Geom. Funct. Anal.*, 21(4):774–819, 2011.

[9] J. B. H. Conway. *A Course in Functional Analysis.* Springer, New York, second edition, 1985.

[10] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Ann. Math.*, 162:439–485, 2005.

[11] Y. Filmus, G. Kindler, N. Lifshitz, and D. Minzer. Hypercontractivity on the symmetric group. Preprint. arXiv:2009.05503, 2020.

[12] W. Fulton and J. Harris. *Representation Theory: A First Course.* Springer, New York, 2013.

[13] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. De Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 516–525, 2007.

[14] W. T. Gowers. Quasirandom Groups. *Combin. Probab. Comput.*, 17:363–387, 2008.

[15] B. Green and I. Z. Ruzsa. Sum-free sets in Abelian groups. *Israel J. Math.*, 147:157–188, 2005.

[16] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97:1061–1083, 1975.

[17] H. A. Helfgott. Growth and generation in $SL_2(\mathbb{Z}/p\mathbb{Z})$. *Ann. Math.*, 167:601–623, 2008.

[18] H. A. Helfgott. Growth in groups: ideas and perspectives. *Bull. Amer. Math. Soc.*, 52(3):357–413, 2015.

[19] H. A. Helfgott and Á. Seress. On the diameter of permutation groups. *Ann. Math.*, 179:611–658, 2014.

[20] R. Henstock and A. M. Macbeath. On the Measure of Sum-Sets. (I) the Theorems of Brunn, Minkowski and Lusternik. *Proc. Lond. Math. Soc.*, 3:182–194, 1953.

[21] J. W. Jenkins. Growth of connected locally compact groups. *J. Funct. Anal.*, 12:113–127, 1973.

[22] T. Jiang. How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.*, 34:1497–1529, 2006.

[23] Y. Jing, C.-M. Tran, and R. Zhang. Effective measure growth for small sets in compact Lie groups. In preparation.

[24] Y. Jing, C.-M. Tran, and R. Zhang. A non-Abelian Brunn-Minkowski inequality. Preprint. arXiv:2101.07782, 2021.

[25] Y. Jing, C.-M. Tran, and R. Zhang. Measure doubling of small sets in $\mathrm{SO}_3(\mathbb{R})$. Preprint. arXiv:2304.09619, 2023.

[26] J. Kahn, G. Kalai, and N. Linial. The influences of variables on Boolean functions. *Proceedings of the 29th Annual Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[27] P. Keevash and N. Lifshitz. Sharp hypercontractivity for symmetric groups and its applications. Preprint. arXiv:2307.15030, 2023.

[28] J. H. B. Kemperman. On products of sets in a locally compact group. *Fund. Math.*, 56:51–68, 1964.

[29] M. Larsen and A. Shalev. Characters of symmetric groups: sharp bounds and applications. *Invent. Math.*, 174:645–687, 2008.

[30] M. Liebeck and A. Shalev. Diameters of finite simple groups: sharp bounds and applications. *Ann. Math.*, 154:383–406, 2001.

[31] E. Mossel. A quantitative Arrow theorem. *Probab. Theory Relat. Fields*, 154:49–88, 2012.

[32] E. Nelson. Construction of quantum fields from Markoff fields. *J. Funct. Anal.*, 12:97–112, 1973.

[33] N. Nikolov and L. Pyber. Product decompositions of quasirandom groups, and a Jordan type theorem. *Europ. J. Combin.*, 13:1063–1077, 2011.

[34] C. Procesi. *Lie groups: an approach through invariants and representations*. Springer, New York, 2007.

[35] L. Pyber and E. Szabó. Growth in finite simple groups of Lie type. *J. Amer. Math. Soc.*, 29:95–146, 2016.

[36] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 245–254, 2008.

[37] O. Regev and B. Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, pages 31–40, 2011.

[38] N. El Samra and R.C. King. Dimensions of irreducible representations of the classical Lie groups. *Journal of Physics A: Mathematical and General*, 12:2317, 1979.

[39] P. Sarnak and X. Xue. Bounds for multiplicities of automorphic representations. *Duke Math. J.*, 64(1):207–227, 1991.

[40] T. Tao. *Expansion in finite simple groups of Lie type*. American Mathematical Society, Providence, Rhode Island, 2015.

[41] H. Weyl. Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare transformationen. II. *Math. Z.*, 24:328–376, 1926.

A Computing dimensions of high degree representations

A.1 The special orthogonal group SO(n)

Lemma 6.7 (Restated) . *Let $n \geqslant 5$. If ρ is an irreducible representation of* SO(n) *of level $d \geqslant n/2$, then*

$$\dim(\rho) \geqslant \exp(n/32).$$

Proof. This also follows from Weyl's original dimension formulae, together with a short computation. First suppose that $n = 2k+1$ is odd. As mentioned above, the equivalence classes of irreducible representations of $\mathrm{SO}(2k+1, \mathbb{R})$ are in an explicit one-to-one correspondence with the partitions λ (of non-negative integers) whose Young diagrams have at most k rows. Weyl's dimension formula states that for any such partition λ, the corresponding irreducible representation ρ_λ of $\mathrm{SO}(2k+1, \mathbb{R})$ has

$$\dim(\rho_\lambda) = \prod_{1 \leqslant i < j \leqslant k} \frac{\lambda_i - \lambda_j + j - i}{j - i} \prod_{1 \leqslant i \leqslant j \leqslant k} \frac{\lambda_i + \lambda_j + 2k + 1 - i - j}{2k + 1 - i - j}.$$

Fix λ a partition of degree $d \in \mathbb{N}$; trivially,

$$\dim(\rho_\lambda) \geqslant \prod_{1 \leqslant i \leqslant j \leqslant k} \frac{\lambda_i + 2k + 1 - i - j}{2k + 1 - i - j} \geqslant \prod_{\substack{1 \leqslant i \leqslant k/2 \\ k/2 \leqslant j \leqslant k}} \left(1 + \frac{\lambda_i}{2k}\right) = \prod_{1 \leqslant i \leqslant k/2} \left(1 + \frac{\lambda_i}{2k}\right)^{k/2}.$$

If $\lambda_1 \geqslant k$, then the previous product is at least $(1 + 1/2)^{k/2} \geqslant \mathsf{exp}(n/16)$ and we are done, so assume that $\lambda_1 < k$ and hence all λ_i's are smaller than k. For all $0 \leqslant \delta < 1/2$ we have $1 + \delta \geqslant e^{\delta/2}$, so the previous product is at least

$$\prod_{1 \leqslant i \leqslant k/2} e^{\frac{\lambda_i}{4k} \cdot \frac{k}{2}} = e^{\frac{1}{8} \sum_{i=1}^{k/2} \lambda_i}.$$

Since $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k$, we get that $\sum_{i=1}^{k/2} \lambda_i \geqslant \frac{1}{2} \sum_{i=1}^{k} \lambda_i = \frac{d}{2}$, so the last expression is at least $e^{d/16}$ and we are done.

The case of even n is very similar. Let $n = 2k$; then the equivalence classes of irreducible representations of $\mathrm{SO}(2k, \mathbb{R})$ are in an explicit correspondence with the partitions λ (of non-negative integers) whose Young diagrams have at most k rows: this correspondence is one-to-one when the number of rows is less than k, but when the number of rows is equal to k, the correspondence is two-to-one (each partition λ with k rows corresponds to two irreducible representations ρ_λ and $\tilde{\rho}_\lambda$ of the same dimension). Writing $c = c(\lambda) = 1$ if λ has less than k rows and $c = c(\lambda) = 1/2$ if λ has exactly k rows, the dimensions of the corresponding irreducible representations are given by the formula

$$\dim(\rho_\lambda) = c \prod_{1 \leqslant i < j \leqslant k} \left(\frac{\lambda_i - \lambda_j + j - i}{j - i}\right) \left(\frac{\lambda_i + \lambda_j + 2k - i - j}{2k - i - j}\right).$$

From now on we can apply exactly the same argument as in the case of n odd; we omit the details. □

A.2 The spin group Spin(n)

Lemma A.1. *Let ρ_λ be as in Theorem 6.9. Then*

$$\dim(\rho_\lambda) = \prod_{1 \leqslant i < j \leqslant k} \frac{\lambda_i - \lambda_j - i + j}{j - i} \prod_{1 \leqslant i \leqslant j \leqslant k} \frac{\lambda_i + \lambda_j + 2k + 1 - i - j}{2k + 1 - i - j} \geqslant 2^{-\Omega(n)}.$$

Proof. Indeed, if a_k is odd then we must have $a_k \geqslant 1$ and therefore $\lambda_i \geqslant 1/2$ for all $i \in [k]$.

Hence, rather crudely, we have

$$\begin{aligned}
\dim(\rho_\lambda) &\geqslant \prod_{1\leqslant i\leqslant j\leqslant k} \frac{\lambda_i+\lambda_j+2k+1-i-j}{2k+1-i-j} \\
&= \prod_{1\leqslant i\leqslant j\leqslant k} \left(1+\frac{\lambda_i+\lambda_j}{2k+1-i-j}\right) \\
&\geqslant \prod_{\substack{1\leqslant i\leqslant k/2,\\ k/2\leqslant j\leqslant k}} \left(1+\frac{\lambda_i}{2k}\right) \\
&\geqslant \prod_{1\leqslant i\leqslant k/2} \left(1+\frac{\lambda_i}{2k}\right)^{k/2} \\
&\geqslant \prod_{1\leqslant i\leqslant k/2} \left(1+\frac{\min\{\lambda_i,2k\}}{2k}\right)^{k/2} \\
&\geqslant \prod_{1\leqslant i\leqslant k/2} \exp(\min\{\lambda_i,2k\}/8) \\
&= \exp\left(\tfrac{1}{8}\sum_{1\leqslant i\leqslant k/2} \min\{\lambda_i,2k\}\right) \\
&\geqslant \exp(k/32) \\
&= \exp((n-1)/64),
\end{aligned}$$

as required, using the fact that $1+x \geqslant e^{x/2}$ for all $x \leqslant 1$.

For all $n = 2k \geqslant 6$ even, we have

$$\dim(\rho_\lambda) = \prod_{1\leqslant i<j\leqslant k} \frac{\lambda_i-\lambda_j-i+j}{j-i}\frac{\lambda_i+\lambda_j+2k-i-j}{2k-i-j},$$

where the k-tuple $\boldsymbol{\lambda} = (\lambda_1,\lambda_2,\ldots\lambda_k)$ ranges over all k-tuples defined by

$$\lambda_i = a_i+a_{i+1}+\ldots+a_{k-2}+\tfrac{1}{2}(a_{k-1}+a_k)\ \forall\, i\leqslant k-2, \quad \lambda_{k-1} = \tfrac{1}{2}(a_{k-1}+a_k), \quad \lambda_k = \tfrac{1}{2}(a_k-a_{k-1}),$$

for some $(a_i)_{i=1}^k \in (\mathbb{N}\cup\{0\})^k$. The case of $a_{k-1}+a_k$ even corresponds to irreducible representations of $\mathrm{Spin}(n)$ that are also irreducible representations of $\mathrm{SO}(n,\mathbb{R})$; the dimensions of these were bounded in the previous subsection. The case of $a_{k-1}+a_k$ odd corresponds to 'new' irreducible representations of $\mathrm{Spin}(n)$, but the above equation quickly implies that any such has dimension at least $2^{\Omega(n)}$. Indeed, if $a_{k-1}+a_k$ is odd then we must have

$a_{k-1} + a_k \geqslant 1$ and therefore $\lambda_i \geqslant 1/2$ for all $i \leqslant k-1$. Hence, again rather crudely, we have

$$\begin{aligned}
\dim(\rho_\lambda) &\geqslant \prod_{1 \leqslant i < j \leqslant k} \frac{\lambda_i + \lambda_j + 2k - i - j}{2k - i - j} \\
&= \prod_{1 \leqslant i < j \leqslant k} \left(1 + \frac{\lambda_i + \lambda_j}{2k - i - j}\right) \\
&\geqslant \prod_{\substack{1 \leqslant i \leqslant k/2, \\ k/2 < j \leqslant k}} \left(1 + \frac{\lambda_i}{2k}\right) \\
&\geqslant \prod_{1 \leqslant i \leqslant k/2} \left(1 + \frac{\lambda_i}{2k}\right)^{(k-1)/2} \\
&\geqslant \prod_{1 \leqslant i \leqslant k/2} \left(1 + \frac{\min\{\lambda_i, 2k\}}{2k}\right)^{k/4} \\
&\geqslant \prod_{1 \leqslant i \leqslant k/2} \exp(\min\{\lambda_i, 2k\}/16) \\
&= \exp\left(\tfrac{1}{16} \sum_{1 \leqslant i \leqslant k/2} \min\{\lambda_i, 2k\}\right) \\
&\geqslant \exp(k/64) \\
&= \exp(n/128),
\end{aligned}$$

as required, again using the fact that $1 + x \geqslant e^{x/2}$ for all $x \leqslant 1$. □

A.3 The special unitary group $\mathbf{SU}(n)$

Lemma 6.26 (Restated) . *For all D, if ρ is an irreducible representation of level D, then* $\mathsf{dim}(\rho) \geqslant 2^{\Omega(\min(D,n))}$.

Proof. The dimension of the irreducible representation corresponding to λ is

$$\prod_{1 \leqslant i < j \leqslant n} \frac{\lambda_i - \lambda_j + j - i}{j - i}.$$

Assume first that $\lambda_1 \geqslant n$. If $\lambda_{\lfloor n/2 \rfloor + 1} \leqslant n/2$ then, considering all the terms in the above product corresponding to $i = 1$ and $j \geqslant \lfloor n/2 \rfloor + 1$, we see that the above product satisfies

$$\prod_{1 \leqslant i < j \leqslant n} \frac{\lambda_i - \lambda_j + j - i}{j - i} \geqslant \prod_{j \geqslant \lfloor n/2 \rfloor + 1} \frac{\lambda_1 - \lambda_j + j - 1}{j - 1}$$

$$\geqslant \prod_{j \geqslant \lfloor n/2 \rfloor + 1} \frac{\lambda_1 - \lambda_j + n - 1}{n - 1}$$

$$\geqslant \prod_{j \geqslant \lfloor n/2 \rfloor + 1} \frac{n - n/2 + n - 1}{n - 1}$$

$$= \exp(\Theta(n)).$$

If, on the other hand, we have $\lambda_{\lfloor n/2 \rfloor + 1} > n/2$, then considering all the terms in the above product corresponding to $j = n$ and $i \leqslant \lfloor n/2 \rfloor + 1$, we obtain

$$\prod_{1 \leqslant i < j \leqslant n} \frac{\lambda_i - \lambda_j + j - i}{j - i} \geqslant \prod_{i \leqslant \lfloor n/2 \rfloor + 1} \frac{\lambda_i - \lambda_n + n - i}{n - i}$$

$$\geqslant \prod_{i \leqslant \lfloor n/2 \rfloor + 1} \frac{\lambda_i - \lambda_n + n - 1}{n - 1}$$

$$\geqslant \prod_{i \leqslant \lfloor n/2 \rfloor + 1} \frac{n/2 - 0 + n - 1}{n - 1}$$

$$= \exp(\Theta(n)).$$

We may henceforth assume that $\lambda_1 < n$. In this case we have $\lambda_i - \lambda_j < n$ for all i, j. Hence, the above product satisfies

$$\begin{aligned}
\prod_{1\leqslant i<j\leqslant n} \frac{\lambda_i - \lambda_j + j - i}{j-i} &\geqslant \prod_{1\leqslant i<j\leqslant n} \frac{\lambda_i - \lambda_j + n - 1}{n-1} \\
&\geqslant \prod_{1\leqslant i<j\leqslant n} \left(1 + \frac{\lambda_i - \lambda_j}{n-1}\right) \\
&\geqslant \prod_{1\leqslant i<j\leqslant n} \exp\left(\frac{\lambda_i - \lambda_j}{2(n-1)}\right) \\
&= \exp\left(\frac{1}{2(n-1)} \sum_{1\leqslant i<j\leqslant n} (\lambda_i - \lambda_j)\right) \\
&= \exp\left(\frac{1}{2(n-1)} \sum_{1\leqslant i<j\leqslant n} (a_i + \ldots + a_{j-1})\right) \\
&= \exp\left(\frac{(n-1)a_1 + 2(n-2)a_2 + \ldots + 2(n-2)a_{n-2} + (n-1)a_{n-1}}{2(n-1)}\right) \\
&\geqslant \exp(\Theta(D)).
\end{aligned}$$

□

B The required adaptations for showing that Sp(n) and SU(n) are good

Here we complete the proof that Sp(n) and SU(n) are good. In fact, we will only explain how to adapt the proof to Sp(n) as SU(n) is only simpler.

To construct our noise operator T_ρ on Sp(n) we use the identification

$$\mathrm{Sp}(n) = \{X \in \mathbb{H}^{n\times n} :\ XX^h = I\} = \{X \in \mathbb{H}^{n\times n} :\ X^hX = I\}$$

of Sp(n) with the group of unitary quaternionic matrices. (Here, X^h denotes the quaternionic conjugate of X, i.e. $(X^h)_{p,q} = \overline{X_{q,p}}$, where $\overline{a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}$ for $a, b, c, d \in \mathbb{R}$.) We couple Sp$(n)$ with the space $(\mathbb{H}^{n\times n}, \gamma)$ of quaternionic normal random matrices, where the real part, the **i**-part (= coefficient of **i**), the **j**-part and the **k**-part of each entry are independent normal (real-valued) random variables with mean zero and variance 1/4, and all the entries are independent. (The following terminology will be useful in the sequel. Recall that a *standard normal quaternionic* or $\mathcal{QN}$ random variable is a quaternion-valued random variable where the real part, the **i**-part, the **j**-part and the **k**-part of the value of the random variable are independent normal (real-valued) random variables with mean zero and variance 1/4; so a quaternionic normal random matrix $\sim (\mathbb{H}^{n\times n}, \gamma)$ is simply a matrix where each entry is an independent $\mathcal{QN}$ random variable. A $\mathcal{QN}$ *random vector in* n *dimensions* is a vector of n independent $\mathcal{QN}$ random variables.)

To make this coupling work, we need to define a 'Gram-Schmidt' type process (on the columns, and also on the rows) which, when applied to a matrix in $(\mathbb{H}^{n\times n},\gamma)$, yields an element of $\mathrm{Sp}(n)$ with probability one. We define two inner products on $\mathbb{H}^n$:

$$\langle x,y\rangle := \sum_{i=1}^{n} \overline{x_i} y_i, \quad \langle x,y\rangle' := \sum_{i=1}^{n} x_i \overline{y_i}.$$

It is easy to check that, for $H\in\mathbb{H}^{n\times n}$, we have $\langle Hx,Hy\rangle = \langle x,y\rangle$ for all $x,y\in\mathbb{H}^n$ if and only if $H\in\mathrm{Sp}(n)$.

Our Gram-Schmidt process on the columns, $\mathsf{GS}_{\mathsf{col}}(X)$ for $X\in\mathbb{H}^{n\times n}$, is defined as follows. If $c_1,c_2,\ldots,c_n$ denote the columns of X, then we (inductively) define

$$\gamma_k := c_k - \sum_{\ell<k} \tilde{c}_\ell \langle \tilde{c}_\ell, c_k\rangle$$

and (if $\gamma_k\neq 0$)

$$\tilde{c}_k := \gamma_k / \sqrt{\langle \gamma_k,\gamma_k\rangle},$$

for $1\leqslant k\leqslant n$. If any $\gamma_k=0$ then $\mathsf{GS}_{\mathsf{col}}(X)$ is undefined; otherwise we define $\mathsf{GS}_{\mathsf{col}}(X)$ to be the matrix in $\mathbb{H}^{n\times n}$ with columns $\tilde{c}_1,\tilde{c}_2,\ldots,\tilde{c}_n$; it is easy to see that this matrix ($\tilde{X}$, say) is an element of $\mathrm{Sp}(n)$. (One checks, by induction on k, that $\langle\tilde{c}_\ell,\tilde{c}_k\rangle = \delta_{j,k}$ for all $1\leqslant j\leqslant k\leqslant n$; taking conjugates this implies that $\langle\tilde{c}_k,\tilde{c}_\ell\rangle=\delta_{j,k}$ for all $1\leqslant j\leqslant k\leqslant n$, and these two statements together imply that $\tilde{X}\in\mathrm{Sp}(n)$. It is clear that, if X is sampled according to γ, then all the γ_k are non-zero with probability one, so $\mathsf{GS}_{\mathsf{col}}(X)$ is defined outside a set of zero probability measure.

We define the Gaussian noise operator $U_\rho : L^2(\mathbb{H}^{n\times n},\gamma)\to L^2(\mathbb{H}^{n\times n},\gamma)$ in the obvious way: for $f\in L^2(\mathbb{H}^{n\times n},\gamma)$ we define

$$U_\rho f(X) = \mathbb{E}_{Y\sim\gamma}[f(\rho X+\sqrt{1-\rho^2}Y)].$$

It is easy to check that U_ρ is self-adjoint; indeed,

$$\mathbb{E}_{X\sim\gamma}[\overline{f(X)}U_\rho g(X)] = \mathbb{E}_{X,Z\sim\gamma,\ \rho\text{-correlated}}\overline{f(X)}g(Z) = \mathbb{E}_{X\sim\gamma}[\overline{U_\rho f(X)}g(X)].$$

As in the case of $\mathrm{SO}(n)$, for $f\in L^2(\mathrm{Sp}(n))$ we define $\mathrm{T}_{\mathsf{col}} : L^2(\mathrm{Sp}(n),\mu)\to L^2(\mathbb{H}^{n\times n},\gamma)$ by

$$\mathrm{T}_{\mathsf{col}}f(X) = f(\mathsf{GS}_{\mathsf{col}}(X)),$$

where μ denotes the Haar measure on $\mathrm{Sp}(n)$. We similarly let T^*_{col} denote its (Hilbert-space) adjoint.

Again, as in the case of $\mathrm{SO}(n)$, we define

$$\mathrm{T}_\rho = \mathbb{E}_{V\sim\mathrm{Sp}(n)}[R_V^*\mathrm{T}^*_{\mathsf{col}}U_\rho\mathrm{T}_{\mathsf{col}}R_V].$$

Since U_ρ is self-adjoint, so is T_ρ. The fact that $\mathrm{T}_{\mathsf{col}}$ commutes with L_V for all $V\in\mathrm{Sp}(n)$ follows from the fact that $\mathsf{GS}_{\mathsf{col}}(VX) = V\mathsf{GS}_{\mathsf{col}}(X)$ for all $X\in\mathbb{H}^{n\times n}$ and all $V\in\mathrm{Sp}(n)$, which in turn follows from the fact that $\langle Vx, Vy\rangle = \langle x, y\rangle$ for all $x, y \in \mathbb{H}^n$ and all $V \in \mathrm{Sp}(n)$. The fact that L_V and R_V both commute with U_ρ follows from the fact that, if $X \sim (\mathbb{H}^{n\times n}, \gamma)$ and $V \in \mathrm{Sp}(n)$, then $VX \sim (\mathbb{H}^{n\times n}, \gamma)$ and $XV \sim (\mathbb{H}^{n\times n}, \gamma)$, as in the case of $\mathrm{SO}(n)$ and $(\mathbb{R}^n, \gamma)$ (the proof is very similar). For any $V \in \mathrm{Sp}(n)$, we have $L_V^* = L_{V^{-1}}$ and $R_V^* = R_{V^{-1}}$, and therefore, taking the adjoints of

$$L_V \mathrm{T}_{\mathsf{col}} = \mathrm{T}_{\mathsf{col}} L_V, \quad R_V \mathrm{T}_{\mathsf{col}} = \mathrm{T}_{\mathsf{col}} R_V$$

yields

$$\mathrm{T}^*_{\mathsf{col}} L_{V^{-1}} = L_{V^{-1}} \mathrm{T}^*_{\mathsf{col}}, \quad \mathrm{T}^*_{\mathsf{col}} R_{V^{-1}} = R_{V^{-1}} \mathrm{T}^*_{\mathsf{col}},$$

and therefore L_V and R_V commute with $\mathrm{T}^*_{\mathsf{col}}$, for all $V \in \mathrm{Sp}(n)$. It follows that L_V commutes with T_ρ. The fact that for any $W \in \mathrm{Sp}(n)$, R_W commutes with T_ρ, follows from a simple change of variables:

$$\begin{aligned}
R_W \mathrm{T}_\rho &= R_W \mathbb{E}_{V\sim\mathrm{Sp}(n)}[R_V^* \mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} R_V] \\
&= \mathbb{E}_{V\sim\mathrm{Sp}(n)}[R^*_{W^{-1}} R_V^* \mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} R_V] \\
&= \mathbb{E}_{V\sim\mathrm{Sp}(n)}[(R_V R_{W^{-1}})^* \mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} R_V] \\
&= \mathbb{E}_{V\sim\mathrm{Sp}(n)}[R_{VW^{-1}} \mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} R_V] \\
&= \mathbb{E}_{V\sim\mathrm{Sp}(n)}[R_V \mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} R_{VW}] \\
&= \mathbb{E}_{V\sim\mathrm{Sp}(n)}[R_V \mathrm{T}^*_{\mathsf{col}} U_\rho \mathrm{T}_{\mathsf{col}} R_V R_W] \\
&= \mathrm{T}_\rho R_W.
\end{aligned}$$

Hence, T_ρ commutes with the action of $\mathrm{Sp}(n)$ from both left and right, as in the $\mathrm{SO}(n)$ case.

We also need to define an analogue of $\mathrm{T}_{\mathsf{row}}$. However, this is a little different to in the $\mathrm{SO}(n)$ case. For $V \in \mathbb{H}^{n\times n}$, $X \in \mathrm{Sp}(n)$ does not imply that $\langle e_i X, e_j X\rangle = \delta_{i,j}$ for each $i, j \in [n]$ (the latter would be the analogue of 'orthonormal rows'). The condition $X \in \mathrm{Sp}(n)$ is, however, equivalent to the condition $\langle e_i X, e_j X\rangle' = \delta_{i,j}$ for each $i, j \in [n]$ (which is in turn equivalent to the condition $\langle e_i \overline{X}, e_j \overline{X}\rangle = \delta_{i,j}$ for each $i, j \in [n]$). We therefore define our Gram-Schmidt process on the rows, $\mathsf{GS}_{\mathsf{row}}(X)$ for $X \in \mathbb{H}^{n\times n}$, as follows. If $r_1, r_2, \ldots, r_n$ denote the rows of X, then we (inductively) define

$$\delta_k := r_k - \sum_{\ell<k} \overline{\langle \tilde{r}_\ell, r_k\rangle'} \tilde{r}_\ell$$

and (if $\delta_k \neq 0$)

$$\tilde{r}_k := \delta_k / \sqrt{\langle \delta_k, \delta_k\rangle'},$$

for $1 \leqslant k \leqslant n$. If any $\delta_k = 0$, then $\mathsf{GS}_{\mathsf{row}}(X)$ is undefined; otherwise we define $\mathsf{GS}_{\mathsf{row}}(X)$ to be the matrix in $\mathbb{H}^{n\times n}$ with rows $\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_n$; it is easy to see that this matrix ($\tilde{X}$, say) is an element of $\mathrm{Sp}(n)$, similarly to in the case of $\mathsf{GS}_{\mathsf{col}}(X)$. Again, as with $\mathsf{GS}_{\mathsf{col}}(X)$, if X is sampled according to γ, then all the δ_k are non-zero with probability one, so $\mathsf{GS}_{\mathsf{row}}(X)$ is defined outside a set of zero probability measure.

For $f \in L^2(\mathrm{Sp}(n))$, we define $\mathrm{T}_{\mathsf{row}} : L^2(\mathrm{Sp}(n), \mu) \to L^2(\mathbb{H}^{n\times n}, \gamma)$ by

$$\mathrm{T}_{\mathsf{row}} f(X) = f(\mathsf{GS}_{\mathsf{row}}(X)),$$

where μ denotes the Haar measure on $\mathrm{Sp}(n)$; of course, we let T^*_{row} denote its (Hilbert-space) adjoint.

As in the $\mathrm{SO}(n)$ case, we observe that $\mathrm{T}_{\mathsf{row}}$ commutes with R_V for all $V \in \mathrm{Sp}(n)$; this follows from the fact that $\mathsf{GS}_{\mathsf{row}}(XV) = \mathsf{GS}_{\mathsf{row}}(X)V$ for all $X \in \mathbb{H}^{n\times n}$ and all $V \in \mathrm{Sp}(n)$, which in turn follows from the fact that $\langle xV, yV\rangle' = \langle x, y\rangle'$ for all $x, y \in \mathbb{H}^n$ and all $V \in \mathrm{Sp}(n)$.

Similarly to as in the $\mathrm{SO}(n)$ case, if $Y \sim (\mathbb{H}^{n\times n}, \gamma)$, and $\sqrt{n}X = \mathsf{GS}_{\mathsf{col}}(Y)$, we obtain $Y = XG$, where $g_{i,i}$ is $1/\sqrt{n}$ times the (Euclidean) length of a $\mathcal{QN}$ random vector in $n-i+1$ dimensions, $g_{i,j} = 0$ for all $i > j$, $g_{i,j}$ is $1/\sqrt{n}$ times a $\mathcal{QN}$ random variable, the entries of G are independent, and independent of all the entries of X. This (distribution over) quaternionic upper-triangular matrices G is our Gaussian Maker Distribution (or GMD, for short) in the $\mathrm{Sp}(n)$ case. To generalise the $\mathrm{SO}(n)$ proof, the only (important) facts we need are the fact that $\mathrm{Sp}(n)$ acts transitively (from either the left or the right) on the set $\mathcal{S} = \{v \in \mathbb{H}^n : \langle v, v\rangle = 1\}$ of quaternionic vectors of unit norm, and that $(\mathbb{H}^{n\times n}, \gamma)$ is invariant under both left and right actions of $\mathrm{Sp}(n)$.

We can therefore write

$$\mathrm{T}^*_{\mathsf{col}} f(X) = \mathbb{E}_{G\sim \mathrm{GMD}} f(XG) \quad \forall X \in \mathrm{Sp}(n).$$

Similarly, we can write

$$\mathrm{T}^*_{\mathsf{row}} f(X) = \mathbb{E}_{G\sim \mathrm{GMD}} f(G^T X) \quad \forall X \in \mathrm{Sp}(n).$$

Our basis of functions on $L^2(\mathbb{H}^{n\times n}, \gamma)$ consists this time of monomials where each variable is a real-part, an $\mathbf{i}$-part, a $\mathbf{j}$-part or a $\mathbf{k}$-part of one of the matrix entries. We say such a monomial is *comfortable* if it is a complex linear combination of monomials in which no two variables come from the same row, no two variables come from the same column, and all variables come from the first $\lfloor n/2 \rfloor$ rows and the first $\lfloor n/2 \rfloor$ columns. Similarly, we say it is *row-comfortable* (respectively *column-comfortable* if it is a complex linear combination of monomials in which no two variables come from the same row (respectively column), and all variables come from the first $\lfloor n/2 \rfloor$ rows (respectively columns). Since the real-part, the $\mathbf{i}$-part, the $\mathbf{j}$-part and the $\mathbf{k}$-part of each matrix entry is $N(0, 1/4)$-distributed rather than $N(0, 1)$-distributed, to guarantee orthonormality we must multiply by a factor of 2^d, so a 'generic' row-comfortable monomial of degree d is of the form

$$2^d (X_{i_1,j_1})_{q_1\text{-part}} \cdot (X_{i_2,j_2})_{q_2\text{-part}} \cdot \ldots \cdot (X_{i_d,j_d})_{q_d\text{-part}},$$

where $q_k \in \{\text{real}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ for all $k \in [d]$ and $i_1, \ldots, i_d$ are distinct integers between 1 and $n/2$; for brevity we denote this by H_S where $S = \{(i_1, j_1; q_1), (i_2, j_2; q_2), \ldots, (i_d, j_d; q_d)\}$.

Our 'standard' comfortable degree-d monomial is:

$$X \mapsto 2^d (X_{1,1})_{\text{real-part}} \cdot (X_{2,2})_{\text{real-part}} \cdot \ldots \cdot (X_{d,d})_{\text{real-part}},$$

denoted $H_{\{(1,1;\mathbb{R}),(2,2;\mathbb{R}),\ldots,(d,d;\mathbb{R})\}} := H_{S_0}$, for brevity.

If $S = \{(1,j_1;q_1),(2,j_2;q_2),\ldots,(d,j_d;q_d)\}$ is such that the j_k are all distinct, then $H_S = R_{V_S} H_{S_0}$, where $V_S = \Sigma D$, Σ is the permutation matrix corresponding to some permutation $\sigma \in S_n$ satisfying $\sigma^{-1}(i) = j_i$ for all $i \in [d]$ ($\Sigma_{i,j} := \delta_{\{j=\sigma(i)\}}$ for all $i,j \in [n]$), and D is a diagonal matrix with $D_{i,i} = \overline{q_i}$ for all $i \in [d]$. It follows that $H_{S_0} = R_{V_S^{-1}} H_S$. Since $V_S \in \mathrm{Sp}(n)$ for any such V_S, and since $V_S V \sim \mathrm{Sp}(n)$ for $V \sim \mathrm{Sp}(n)$, we have

$$\begin{aligned}
\mathbb{E}_{V\sim\mathrm{Sp}(n)} \|\Pi_{\mathsf{comf},d} R_V f\|^2_{L^2(\gamma)} &= \sum_{S\colon H_S \text{ comfortable}} \mathbb{E}_{V\sim\mathrm{Sp}(n)} |\langle R_V f, H_S\rangle|^2 \\
&= \sum_{S\colon H_S \text{ comfortable}} \mathbb{E}_{V\sim\mathrm{Sp}(n)} |\langle R_{V_S V} f, H_S\rangle|^2 \\
&= \sum_{S\colon H_S \text{ comfortable}} \mathbb{E}_{V\sim\mathrm{Sp}(n)} |\langle R_{V_S} R_V f, H_S\rangle|^2 \\
&= \sum_{S\colon H_S \text{ comfortable}} \mathbb{E}_{V_S V\sim\mathrm{Sp}(n)} |\langle R_V f, R_{V_S^{-1}} H_S\rangle|^2 \\
&= 4^d \frac{(n/2)!}{((n/2)-d)!} \mathbb{E}_{V\sim\mathrm{Sp}(n)} |\langle R_V f, H_{S_0}\rangle|^2.
\end{aligned}$$

Now suppose that f is a row-comfortable polynomial of pure degree d. Write

$$f = \sum_{S=\{(1,j_1;q_1),\ldots,(d,j_d;q_d)\}} \hat{f}(S) H_S.$$

We have

$$R_V H_S(X) = H_S(XV) = 2^d \prod_{k=1}^{d} ((XV)_{k,j_k})_{q_k\text{-part}} = 2^d \prod_{k=1}^{d} \left(\sum_{i=1}^{n} X_{k,i} V_{i,j_k} \right)_{q_k\text{-part}},$$

and therefore

$$\langle R_V H_S(X), H_{S_0}(X)\rangle = \prod_{k=1}^{d} (V_{k,j_k})_{q_k\text{-part}} = 2^{-d} H_S(V).$$

It follows that

$$\langle R_V f, H_{S_0}\rangle = 2^{-d} \sum_S \hat{f}(S) H_S(V) = 2^{-d} f(V),$$

and therefore

$$\begin{aligned}
\mathbb{E}_{V\sim\mathrm{Sp}(n)} \|\Pi_{\mathsf{comf},d} R_V f\|^2_{L^2(\gamma)} &= 4^d \frac{(n/2)!}{((n/2)-d)!} \mathbb{E}_{V\sim\mathrm{Sp}(n)} 4^{-d} |f(V)|^2 \\
&= \frac{(n/2)!}{n^d((n/2)-d)!} \|f\|^2_{L^2(\mu)}.
\end{aligned}$$

where μ is the Haar measure on $\mathrm{Sp}(n)$ dialated by $\sqrt{n}$, which is a measure over $\sqrt{n}\mathrm{Sp}(n)$.

The proof (and statement) of Claim 9.6 is readily adapted. If

$$S = \{(i_1, j_1; q_1), (i_2, j_2; q_2), \dots, (i_d, j_d; q_d)\}$$

is such that the j_k are all distinct, then we have

$$\begin{aligned}
\mathrm{T}^*_{\mathsf{col}} f(X) &= \mathbb{E}_{G\sim\mathrm{GMD}} H_S(XG) \\
&= 2^d \mathbb{E}_{G\sim\mathrm{GMD}} \left[\prod_{k=1}^{d} ((XG)_{i_k,j_k})_{q_k\text{-part}} \right] \\
&= 2^d \mathbb{E}_{G\sim\mathrm{GMD}} \left[\prod_{k=1}^{d} \left(\sum_{\ell=1}^{j_k} (X_{i_k,\ell} G_{\ell,j_k})_{q_k\text{-part}} \right) \right] \\
&= 2^d \prod_{k=1}^{d} \mathbb{E}_{G\sim\mathrm{GMD}} \left[\sum_{\ell=1}^{j_k} (X_{i_k,\ell} G_{\ell,j_k})_{q_k\text{-part}} \right] \\
&= 2^d \prod_{k=1}^{d} \left(\sum_{\ell=1}^{j_k} (X_{i_k,\ell} \mathbb{E}_G[G_{\ell,j_k}])_{q_k\text{-part}} \right) \\
&= 2^d \prod_{k=1}^{d} (X_{i_k,j_k} \mathbb{E}_G[G_{j_k,j_k}])_{q_k\text{-part}} \\
&= 2^d \prod_{k=1}^{d} (X_{i_k,j_k})_{q_k\text{-part}} \mathbb{E}_G[G_{j_k,j_k}] \\
&= \lambda_S H_S(X),
\end{aligned}$$

where

$$\lambda_S := \prod_{k=1}^{d} \mathbb{E}_G[G_{j_k,j_k}].$$

(Note that we used the fact that $\mathbb{E}_G[G_{\ell,j_k}] = 0$ for all $\ell \neq j_k$, and that $\mathbb{E}_G[G_{j_k,j_k}] \in \mathbb{R}$.) The rest of the proof is almost exactly the same as before.

We now need analogues of the lemmas of Appendix C. As in the $\mathrm{O}(n)$ case, we define the 'over-Gaussian' distribution ν to be the distribution of YG, where $Y \sim \gamma$ and $G \sim \mathrm{GMD}$. Fix a comfortable d-junta f and write

$$f = \sum_S a_I (2^d x_S),$$

where for $S = \{(1, i_1; q_1), \dots, (d, i_d; q_d)\}$ (with $i_1, \dots, i_d \in [n]$ distinct, and $q_1, \dots, q_d \in \{\mathrm{real}, \mathbf{i}, \mathbf{j}, \mathbf{k}\} := \mathcal{R}$), we write

$$x_S := \prod_{h=1}^{d} (x_{h,i_h})_{q_h\text{-part}}.$$

Note that $\{2^d x_S\}_S$ forms an orthonormal set of vectors in $L^2(\gamma)$, where S ranges over tuples of the above form.

We need the following analogue of Claim C.1.

Claim B.1. *Let* $S = \{(1, i_1; q_1), \ldots, (d, i_d; q_d)\}$ *where* $i_1, \ldots, i_d \in [n]$ *are distinct and* $q_1, \ldots, q_d \in \{\textit{real}, \mathbf{i}, \mathbf{j}, \mathbf{k}\} := \mathcal{R}$*, and let* $x_S := \prod_{h=1}^{d} (x_{h,i_h})_{q_h\text{-part}}$*. Then*

$$\|2^d x_S\|_{L^2(\nu)}^2 = 1.$$

Proof. In what follows, for $q \in \{\text{real}, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ and $h \in \mathbb{H}$, we define $(h)_{-q\text{-part}} := -(h)_{q\text{-part}}$, for notational convenience. Observe that

$$\|2^d x_S\|_{L^2(\nu)}^2 = 4^d \mathbb{E}_{G\sim\text{GMD}} \mathbb{E}_{Y\sim\gamma} [(((YG)_{1,i_1})_{q_1\text{-part}})^2 (((YG)_{2,i_2})_{q_2\text{-part}})^2 \cdots (((YG)_{d,i_d}))_{q_d\text{-part}}^2].$$

Since for each $h \in [d]$, $((YG)_{h,i_h})_{q_h\text{-part}} = \sum_{k=1}^{i_h} \sum_{r\in\mathcal{R}} (Y_{h,k})_{r\text{-part}} (G_{k,i_h})_{r^{-1}q_h\text{-part}}$ involves only entries of Y in row h and entries of G in column i_h (and the i_h are distinct), the random variables $\{(((YG)_{h,i_h})_{q_h\text{-part}})^2 : \ h \in [d]\}$ form a system of independent random variables, and therefore

$$\|2^d x_S\|_{L^2(\nu)}^2 = 4^d \prod_{h=1}^{d} \mathbb{E}_{G\sim\text{GMD}} \mathbb{E}_{Y\sim\gamma} [(((YG)_{h,i_h})_{q_h\text{-part}})^2].$$

For each $h \in [d]$, we have

$$\begin{aligned}
&\mathbb{E}_{G\sim\text{GMD}} \mathbb{E}_{Y\sim\gamma} [(((YG)_{h,i_h})_{q_h\text{-part}})^2] \\
&= \mathbb{E}_G \mathbb{E}_Y \left[\left(\sum_{k=1}^{i_h} \sum_{r\in\mathcal{R}} (Y_{h,k})_{r\text{-part}} (G_{k,i_h})_{r^{-1}q_h\text{-part}} \right)^2 \right] \\
&= \sum_{(k,r)\neq(k',r')} \mathbb{E}_G \mathbb{E}_Y [(Y_{h,k})_{r\text{-part}} (Y_{h,k'})_{r'\text{-part}} (G_{k,i_h})_{r^{-1}q_h\text{-part}} (G_{k',i_h})_{(r')^{-1}q_h\text{-part}}] \\
&+ \sum_{k=1}^{i_h} \mathbb{E}_{G\sim\text{GMD}} \mathbb{E}_{Y\sim\gamma} [Y_{h,k}^2 G_{k,i_h}^2] \\
&= 0 + \sum_{k=1}^{i_h} \sum_{r\in\mathcal{R}} \mathbb{E}_{G\sim\text{GMD}} [((G_{k,i_h})_{r^{-1}q_h\text{-part}})^2] \mathbb{E}_{Y\sim\gamma} [((Y_{h,k})_{r\text{-part}})^2] \\
&= \sum_{k=1}^{i_h} \sum_{r\in\mathcal{R}} \mathbb{E}_{G\sim\text{GMD}} [((G_{k,i_h})_{r^{-1}q_h\text{-part}})^2] \cdot \tfrac{1}{4} \\
&= \tfrac{1}{4} (4(i_h - 1)(1/(4n)) + (n - i_h + 1)/n) \\
&= 1/4.
\end{aligned}$$

(Here, for the third equality we use the independence of

$$(Y_{h,k})_{r\text{-part}}, (Y_{h,k'})_{r'\text{-part}}, (G_{k,i_h})_{r^{-1}q_h\text{-part}}, (G_{k',i_h})_{(r')^{-1}q_h\text{-part}}$$

and the fact that $(Y_{h,k})_{r\text{-part}}$ and $(Y_{h,k'})_{r'\text{-part}}$ both have zero expectation.) Hence, $\|2^d x_S\|^2_{L^2(\nu)} = 1$, as required. $\square$

Similarly, we need the following analogue of Claim C.2. For $S = \{(1, i_1; q_1), \ldots, (d, i_d; q_d)\}$ and $T = \{(1, j_1; p_1), \ldots, (d, j_d; p_d)\}$ we set

$$d(S,T) := |\{h \in [d] : \ i_h \neq j_h \text{ or } q_h \neq p_h\}|.$$

Claim B.2. *For any S, T such that $d(S,T) = \ell$, we have $\left|\langle 2^d x_S, 2^d x_T\rangle\right|_{L^2(\nu)} \leqslant \varepsilon_\ell$, where*

$$\varepsilon_\ell := 2^{\ell+4} n^{-\ell/2} 2^{d\ell/\sqrt{n}}.$$

To prove this we first need the following simple analogue of Claim C.3.

Claim B.3. *Let $(i_1, \ldots, i_d) \in [n]^d$ and $(j_1, \ldots, j_d) \in [n]^d$ be such that $|\{h \in [d] : \ i_h \neq j_h\}| = \ell$ and such that in the product*

$$(G_{i_1 j_1})_{r_1\text{-part}} (G_{i_2 j_2})_{r_2\text{-part}} \cdots (G_{i_d j_d})_{r_d\text{-part}},$$

no matrix entry of G appears more than twice. Then

$$|\mathbb{E}_{G \sim GMD}\left[(G_{i_1 j_1})_{r_1\text{-part}} (G_{i_2 j_2})_{r_2\text{-part}} \cdots (G_{i_d j_d})_{r_d\text{-part}}\right]| \leqslant \left(\frac{1}{4n}\right)^{\ell/2}.$$

Proof. If, in the product

$$(G_{i_1 j_1})_{r_1\text{-part}} (G_{i_2 j_2})_{r_2\text{-part}} \cdots (G_{i_d j_d})_{r_d\text{-part}},$$

some off-diagonal matrix entry of G appears exactly once, then the expectation of the product is zero. We may therefore assume that every matrix entry of G appears either exactly twice, or not at all, in the above product. If there are exactly ℓ values of h such that $i_h \neq j_h$, then the above expectation factorises into a product of the expectations of the squares of $\ell/2$ off-diagonal and of the squares of $(d-\ell)/2$ diagonal entries:

$$\prod_{k \in \mathcal{D}} \mathbb{E}[((G_{k,k})_{q_k\text{-part}})^2] \prod_{(i,j) \in \mathcal{E}} \mathbb{E}[((G_{i,j})_{r_i\text{-part}})^2],$$

where $\mathcal{E} \subset [n]^2 \setminus \{(k,k) : \ k \in [n]\}$, $|\mathcal{D}| = (d-\ell)/2$, $|\mathcal{E}| = \ell/2$ and $q_k, r_i \in \mathcal{R}$ for all i and k. We have $\mathbb{E}[((G_{i,j})_{r_i\text{-part}})^2] = 1/(4n)$ for all $(i,j) \in \mathcal{E}$ and $\mathbb{E}[((G_{k,k})_{q_k\text{-part}})^2] \leqslant 1$ for all $k \in \mathcal{D}$, proving the claim. $\square$

Proof. Let $\ell \geqslant 1$ and fix $S = \{(1, i_1; q_1), \ldots, (d, i_d; q_d)\}$ and $T = \{(1, j_1; p_1), \ldots, (d, j_d; p_d)\}$ such that $d(S,T) = \ell \geqslant 1$. Since G is upper-triangular and $i_h, j_h \leqslant d$ for all $h \in [d]$, we have

$$((YG)_{h,i_h})_{q_h\text{-part}} = \sum_{k=1}^{i_h} \sum_{r \in \mathcal{R}} (Y_{h,k})_{r\text{-part}} (G_{k,i_h})_{r^{-1}q_h\text{-part}} = \sum_{k=1}^{d} \sum_{r \in \mathcal{R}} (Y_{h,k})_{r\text{-part}} (G_{k,i_h})_{r^{-1}q_h\text{-part}}$$

and

$$((YG)_{h,j_h})_{p_h\text{-part}} = \sum_{k=1}^{j_h} \sum_{r\in\mathcal{R}} (Y_{h,k})_{r\text{-part}} (G_{k,j_h})_{r^{-1}p_h\text{-part}} = \sum_{k=1}^{d} \sum_{r\in\mathcal{R}} (Y_{h,k})_{r\text{-part}} (G_{k,j_h})_{r^{-1}p_h\text{-part}}$$

for all $h \in [d]$. Hence,

$$\begin{aligned} x_S\,(YG) &= \prod_{h=1}^{d} ((YG)_{h,i_h})_{q_h\text{-part}} \\ &= \sum_{\substack{K=(k_1,\dots,k_d)\in[d]^d, \\ R=(r_1,\dots,r_d)\in\mathcal{R}^d}} (Y_{1,k_1})_{r_1\text{-part}} \cdots (Y_{d,k_d})_{r_d\text{-part}} (G_{k_1,i_1})_{r_1^{-1}q_1\text{-part}} \cdots (G_{k_d,i_d})_{r_d^{-1}q_d\text{-part}} \end{aligned}$$

and

$$x_T\,(YG) = \sum_{\substack{K=(k_1,\dots,k_d)\in[d]^d, \\ R=(r_1,\dots,r_d)\in\mathcal{R}^d}} (Y_{1,k_1})_{r_1\text{-part}} \cdots (Y_{d,k_d})_{r_d\text{-part}} (G_{k_1,j_1})_{r_1^{-1}p_1\text{-part}} \cdots (G_{k_d,j_d})_{r_d^{-1}p_d\text{-part}},$$

so, using the fact that, under ν, the $((Y_{i,j})_{r\text{-part}} : i,j \in [n],\ r \in \mathcal{R})$ are independent and of expectation zero (and are independent of the $G_{i,j}$), we obtain

$$\begin{aligned} &\left\langle 2^d x_S, 2^d x_T \right\rangle_\nu \\ &= \sum_{\substack{K\in[d]^d, \\ R\in\mathcal{R}^d}} \mathbb{E}_{G\sim\text{GMD}} \left[(G_{k_1i_1})_{r_1^{-1}q_1\text{-part}} (G_{k_1j_1})_{r_1^{-1}p_1\text{-part}} \cdots (G_{k_di_d})_{r_d^{-1}q_d\text{-part}} (G_{k_dj_d})_{r_d^{-1}p_d\text{-part}} \right]. \end{aligned} \tag{12}$$

For a d-tuple $(K;R) = (k_1;r_1\dots,k_d;r_d) \in ([d]\times\mathcal{R})^d$, we write $m_1 = m_1(K) = m_1(K;R) := |\{h \in [d] :\ j_h = i_h,\ k_h \neq i_h\}|$, $m_2 = m_2(K) = m_2(K;R) := |\{h \in [d] :\ j_h \neq i_h,\ k_h \notin \{i_h, j_h\}\}|$ and $m_3 = m_3(K;R) := |\{h \in [d] :\ j_h \neq i_h,\ k_h \in \{i_h, j_h\}\}|$; note that these quantities depend only on K and not on R. We let $\mathcal{K}\,(m_1,m_2,m_3)$ denote the set of d-tuples $(K;R)$ with parameters m_1, m_2 and m_3. For $(K;R) \in \mathcal{K}\,(m_1,m_2,m_3)$, by Claim 11.3 we have

$$\left| \mathbb{E}_G \left[(G_{k_1i_1})_{r_1^{-1}q_1\text{-part}} (G_{k_1j_1})_{r_1^{-1}p_1\text{-part}} \cdots (G_{k_di_d})_{r_d^{-1}q_d\text{-part}} (G_{k_dj_d})_{r_d^{-1}p_d\text{-part}} \right] \right| \leqslant (4n)^{-\frac{2m_1+2m_2+m_3}{2}}.$$

We further note that, for $(K;R) \in \mathcal{K}(m_1,m_2,m_3)$, the expectation in the above inequality is zero unless the following four conditions hold:

- Whenever $i_h = j_h$ and $k_h \neq i_h$, we have $p_h = q_h$.
- Whenever $i_h = j_h = k_h$ we have $p_h = q_h$ and $r_h^{-1}p_h = \text{real}$.
- Whenever $i_h \neq j_h$ and $k_h = i_h$ we have $r_h^{-1}q_h = \text{real}$.

- Whenever $i_h \neq j_j$ and $k_h = j_h$ we have $r_h^{-1} p_h = \text{real}$.

In view of this we let $\mathcal{K}^*(m_1, m_2, m_3)$ be the set of all d-tuples $(K; R) \in \mathcal{K}(m_1, m_2, m_3)$ such that the above four conditions hold. For $(K; R) \in \mathcal{K}^*(m_1, m_2, m_3)$ we have $m_2 + m_3 = \ell$.

Summing over all K, we see that $|\langle x_S, x_T \rangle|$ is at most

$$\begin{aligned} &\sum_{m_1, m_2, m_3} \sum_{K \in \mathcal{K}^*(m_1, m_2, m_3)} (4n)^{-\frac{2m_1 + 2m_2 + m_3}{2}} \\ &\leqslant \sum_{m_1, m_2, m_3} (4n)^{-\frac{2m_1 + 2m_2 + m_3}{2}} \left| \mathcal{K}^* \left(m_1, m_2, m_3 \right) \right|. \end{aligned}$$

Now

$$\left| \mathcal{K}^* \left(m_1, m_2, m_3 \right) \right| \leqslant \binom{d}{m_1} d^{m_1} \binom{\ell}{m_2} d^{m_2} 2^{m_3} \cdot 4^{m_1 + m_2} \leqslant \frac{d^{2m_1 + m_2} \ell^{m_2} 2^{m_3}}{m_1! m_2!} \cdot 4^{m_1 + m_2};$$

note that the only difference with the corresponding expression in the proof of Claim C.3 is the extra factor of $4^{m_1 + m_2}$, which comes from the fact that r_h can vary freely over $\mathcal{R}$ (and still satisfy the above conditions) when $i_h \neq j_h$ and $k_h \notin \{i_h, j_h\}$, or when $i_h = j_h$ and $k_h \neq i_h$, but in no other cases.

Summing over all m_1, m_2, m_3 with $m_2 + m_3 = \ell$ completes the proof, just as in the proof of Claim C.2; the extra factor of $4^{-\frac{2m_1 + 2m_2 + m_3}{2}}$ cancels out (or more than cancels out) the extra factor of $4^{m_1 + m_2}$. □

The analogue of Lemma 10.9 is proven from the above claims in a very similar way. Writing

$$f = \sum_S \alpha_S (2^d x_S),$$

we obtain

$$\begin{aligned} \|f\|^2_{L^2(\nu)} &\leqslant \sum_S |\alpha_S|^2 \|2^d x_S\|^2_{L^2(\nu)} + \sum_{S \neq T} |\alpha_S| |\alpha_T| \left| \left\langle 2^d x_S, 2^d x_T \right\rangle_\nu \right| \\ &\leqslant \|f\|^2_{L^2(\gamma)} + \sum_{S \neq T} \frac{|a_S|^2 + |a_T|^2}{2} \left| \left\langle 2^d x_S, 2^d x_T \right\rangle_\nu \right| \\ &\leqslant \|f\|^2_{L^2(\gamma)} + \sum_{\ell=1}^d \sum_S |a_S|^2 \left| \{T : d(T, S) = \ell\} \right| \cdot \varepsilon_\ell \\ &\leqslant \|f\|^2_{L^2(\gamma)} + \|f\|^2_{L^2(\gamma)} \sum_{\ell=1}^d \varepsilon_\ell \binom{d}{\ell} \ell! 4^\ell \\ &= \|f\|^2_{L^2(\gamma)} \left(1 + \sum_{\ell=1}^d \varepsilon_\ell (4d)^\ell \right), \end{aligned}$$

and the rest of the proof is essentially unchanged, up to reducing the value of δ by a constant factor.

C Upper bounding the eigenvalues of the Laplacian

C.1 Lower bounding the eigenvalues of the Laplacian in SO(n)

In this section, we prove the following.

Lemma C.1. *Let $\rho \in \hat{\mathrm{SO}}(n)$ be of level D, then the corresponding eigenvalue $-\lambda_\rho$ of the Laplacian Δ satisfies*

$$\lambda_\rho \leqslant C(nD + D^2).$$

To do so, we will use [7, Theorems 2.3, 2.4], which show that the coefficients of the irreducible representations of SO(n) are eigenvectors of the Laplace–Beltrami operator. Furthermore, they establish a 1-to-1 correspondence of these irreducible representation and a system of fundamental weights of SO(n), and give a formula for the eigenvalues of the eigenvectors in the language of fundamental weights. We summarize this discussion with the following result that combines the two results from [7].

Theorem C.2. *Let G be a simply connected Lie group of rank k. Then there are vectors $w_1, \dots, w_k$ such that there is a 1-to-1 correspondence between equivalence classes of irreducible representations of G and the cone*

$$\left\{ \sum_{i=1}^{k} r_i w_i \;\middle|\; r_1, \dots, r_k \in \mathbb{N} \right\}.$$

Furthermore, denoting $\rho = \sum_{i=1}^{k} w_i$, the eigenvalue the entries of the representation corresponding to $v = \sum_{i=1}^{k} r_i w_i$ is $-\|v + \rho\|_2^2 + \|\rho\|_2^2$.

We will choose a known system of fundamental weights of SO(n) as in [34, Section 5.1], and throughout we denote by $E_{i,j} \in \mathbb{R}^{n\times n}$ the diagonal matrix which is 1 on entry i, j and everywhere else is 0; we omit n from notation as it will always be clear from context. The system of fundamental weights depends on whether n is even or odd, and we inspect each case separately.

The case of odd n

Let $n = 2k+1$. In this case, the rank of SO(n) is k, and a system of fundamental weights can be taken as $w_i = \sum_{j=1}^{i} u_j$ for $i \leqslant k-1$ and $w_k = \frac{1}{2}\sum_{j=1}^{k} u_j$, where we have $u_j = E_{j,k+j} - E_{k+j,j}$.

Then, the the equivalence class of representations corresponding to $v = \sum_{i=1}^{k} r_i w_i$ correspond to Young diagrams $\lambda = (\lambda_1, \dots, \lambda_k)$ with k rows where $r_i = \lambda_i - \lambda_{i+1}$ for $i \leqslant k - 1$

and $r_k = 2\lambda_k$, thus the corresponding degree is

$$D = \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k-1} \left(\sum_{i \leqslant j \leqslant k-1} r_j + \frac{r_k}{2} \right) + \frac{r_k}{2} = \sum_{1 \leqslant j \leqslant k-1} j r_j + \frac{k}{2} r_k. \tag{13}$$

We will use this equality now to estimate the eigenvalue as given in Theorem C.2. Using the formula therein, the corresponding eigenvalue is

$$\lambda_v = - \|v + \rho\|_2^2 + \|\rho\|_2^2 ,$$

where here and throughout we think of the matrices v and ρ as n^2 length vector by the natural flattening (equivalently $\langle A, B \rangle = \mathsf{tr}(A^t B)$). Then

$$\lambda_v = -2\langle v, \rho \rangle - \|v\|_2^2 ,$$

and we estimate the norm of v and the inner product between v and ρ. To compute the norm of v we write

$$v = \sum_{i=1}^{k} w_i = \sum_{i=1}^{k-1} r_i \sum_{j=1}^{i} u_j + \frac{r_k}{2} \sum_{j=1}^{k} u_j = \sum_{j=1}^{k} \left(\frac{r_k}{2} + \sum_{i=j}^{k-1} r_i \right) u_j,$$

and since the u_j's are mutually orthogonal and each has 2-norm-squared equal to 2, we get that

$$\|v\|_2^2 = 2 \sum_{j=1}^{k} \left(\frac{r_k}{2} + \sum_{i=j}^{k-1} r_i \right)^2 \leqslant 2 \left(\sum_{j=1}^{k} \left(\frac{r_k}{2} + \sum_{i=j}^{k-1} r_i \right) \right)^2 = 2 \left(\sum_{j=1}^{k} \frac{k r_k}{2} + \sum_{i=1}^{k-1} i r_i \right)^2 ,$$

which is at most $2D^2$ by (13).

To estimate the inner product between v and ρ note that ρ is the vector v in which we take all r_i's to be 1, hence by the computation above

$$\rho = \sum_{i=1}^{k} w_i = \sum_{j=1}^{k} \left(k - j + \frac{1}{2} \right) u_j,$$

and so

$$\langle \rho, v \rangle = 2 \sum_{j=1}^{k} \left(k - j + \frac{1}{2} \right) \left(\frac{r_k}{2} + \sum_{i=j}^{k-1} r_i \right) \leqslant 2k \sum_{j=1}^{k} \left(\frac{r_k}{2} + \sum_{i=j}^{k-1} r_i \right) \leqslant 2kD.$$

where in the last inequality we used (13). Overall, we get that the eigenvalue λ_v satisfies $\lambda_v \geqslant -2D^2 - nD$, as required.

The case of even n

Let $n = 2k$. In this case, the rank of $\mathrm{SO}(n)$ is k, and a system of fundamental weights can be taken as $w_i = \sum_{j=1}^{i} u_j$ for $i \leqslant k-2$, $w_{k-1} = \frac{1}{2}\sum_{j=1}^{k} u_j$ and $w_k = w_{k-1} - u_k$, where again we have $u_j = E_{j,k+j} - E_{k+j,j}$.

Consider the the equivalence class of representations corresponding to $v = \sum_{i=1}^{k} r_i w_i$. We now need to inspect the corresponding Young diagram to relate the r_i's to the degree of the representation, and we recall that the corresponding Young diagram $\lambda = (\lambda_1, \ldots, \lambda_k)$ maybe either have k non-zero rows or at most $k-1$ non-zero rows.

Young diagrams with at most $k-1$ rows. In this case we have that $r_i = \lambda_i - \lambda_{i+1}$ for $i = 1, \ldots, k$, and so the degree is

$$D = \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k}\sum_{j=i}^{k-1} r_j = \sum_{j=1}^{k-1} j r_j.$$

Using the same estimates as before, we get that

$$v = \sum_{i=1}^{k-2} r_i \sum_{j=1}^{i} u_j + \sum_{j=1}^{k} \frac{r_{k-1}}{2} u_j = \sum_{j=1}^{k}\left(\sum_{i=j}^{k-2} r_i + \frac{r_{k-1}}{2}\right) u_j,$$

so

$$\|v\|_2^2 \leqslant 2\sum_{j=1}^{k}\left(\sum_{i=j}^{k-2} r_i + \frac{r_{k-1}}{2}\right)^2 \leqslant 2\left(\sum_{i=j}^{k-2} i r_i + \frac{k r_{k-1}}{2}\right)^2 \leqslant 2D^2.$$

Also, we get that $\rho = \sum_{j=1}^{k-2}(k - \frac{1}{2} - j)u_j + \frac{1}{2}u_{k-1} + \frac{1}{2}u_k$ and so

$$\langle v, \rho\rangle = 2\sum_{j=1}^{k-2}\left(k - \frac{1}{2} - j\right)\left(\sum_{i=j}^{k-2} r_i + \frac{r_{k-1}}{2}\right) + r_{k-1} \leqslant 2k\sum_{j=1}^{k-2}\left(\sum_{i=j}^{k-2} r_i + \frac{r_{k-1}}{2}\right) + r_{k-1} \leqslant 2kD + D,$$

which is at most $2nD$.

Young diagrams with k rows. In this case we have that $r_i = \lambda_i - \lambda_{i+1}$ for $i = 1, \ldots, k-2$, and (r_{k-1}, r_k) are either $(\lambda_{k-1} - \lambda_k, \lambda_{k-1} + \lambda_k)$ or $(\lambda_{k-1} + \lambda_k, \lambda_{k-1} - \lambda_k)$. The computation in both cases is similar and goes along the same lines as the computations so far, hence we focus on $(r_{k-1}, r_k) = (\lambda_{k-1} - \lambda_k, \lambda_{k-1} + \lambda_k)$ for the sake of concreteness.

The degree is

$$D = \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k-2}\left(\sum_{j=i}^{k-2} r_j + \frac{1}{2}(r_{k-1} + r_k)\right) = \sum_{j=1}^{k-2} j r_j + \frac{k-2}{2} r_{k-1} + \frac{k-2}{2} r_k.$$

We have the same formula for v and ρ as before, and so

$$\|v\|_2^2 \leqslant 2\sum_{j=1}^{k}\left(\sum_{i=j}^{k-2} r_i + \frac{r_{k-1}}{2}\right)^2 \leqslant 2\left(\sum_{i=j}^{k-2} i r_i + \frac{k r_{k-1}}{2}\right)^2 \leqslant 2D^2,$$

as well as

$$\langle v,\rho\rangle = 2\sum_{j=1}^{k-2}\left(k-\frac{1}{2}-j\right)\left(\sum_{i=j}^{k-2} r_i + \frac{r_{k-1}}{2}\right)+r_{k-1} \leqslant 2k\sum_{j=1}^{k-2}\left(\sum_{i=j}^{k-2} r_i + \frac{r_{k-1}}{2}\right)+r_{k-1} \leqslant 2kD+D,$$

which is at most $2nD$.

C.2 The eigenvalues of the Laplace–Beltrami operator in SU(n)

The following lemma is analogous to Lemma C.1 and gives abound on the eigenvalues of the Laplace–Beltrami operator of SU(n).

Lemma C.3. *For $\rho \in \widehat{\mathrm{SU}(n)}$ of level D, the corresponding eigenvalue $-\lambda_\rho$ of Δ satisfies*

$$\lambda_\rho \leqslant C(nD + D^2),$$

where C is an absolute constant.

Similarly to in the case of SO(n), using the formulae in [7] (multiplying by a factor of $2n$, similarly to before) and the systems of fundamental weights in [34], one can show that for all $\rho \in \widehat{SU(n)}$ of level D, the corresponding eigenvalue $-\lambda_\rho$ of Δ satisfies

$$\lambda_\rho \leqslant C(nD + D^2),$$

where $C > 0$ is an absolute constant. We use the system of fundamental weights

$$w_i = \sum_{j=1}^{i} e_j - \tfrac{i}{n}\sum_{j=1}^{n} e_j \ (1 \leqslant i \leqslant n-1),$$

where $\{e_i\}_{i=1}^n$ is the standard basis of $\mathbb{R}^n$; here, e_i corresponds to

$$\mathbf{i}E_{i,i}, \tag{14}$$

where $\mathbf{i} = \sqrt{-1}$ and $E_{i,j}$ is the matrix with a 1 in the (i,j)th-entry and zeros elsewhere. For each $1 \leqslant k \leqslant l \leqslant n-1$ we have

$$\langle w_k, w_l\rangle = k(n-l)/n.$$

Set $\sigma := \sum_{i=1}^{n-1} w_i$. For a partition λ whose Young diagram has less than n rows, the corresponding weight vector is

$$\mathbf{v} = \sum_{i=1}^{n-1} a_i w_i,$$

where $a_i = \lambda_i - \lambda_{i+1}$ for all $i \in [n-1]$ and $\lambda_n := 0$; the level D of the corresponding representation is given by

$$D = \sum_{i=1}^{n-1} a_i \min\{i, n-i\}.$$

It follows that, if $\mathbf{v} = \sum_{k=1}^{n-1} a_k w_k$, then

$$\begin{aligned}
\langle \mathbf{v}, \sigma \rangle &= \left\langle \sum_{k=1}^{n-1} a_k w_k, \sum_{k=1}^{n-1} w_k \right\rangle \\
&= \sum_{1 \leqslant k \leqslant l \leqslant n-1} a_k k(n-l)/n + \sum_{1 \leqslant l < k \leqslant n-1} a_k l(n-k)/n \\
&= \sum_{k=1}^{n-1} (n-k)(n-k+1) k a_k/(2n) + \sum_{k=1}^{n-1} k(k+1) l(n-k) a_k/(2n) \\
&\leqslant nD,
\end{aligned}$$

and

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{v} \rangle &= \left\langle \sum_{k=1}^{n-1} a_k w_k, \sum_{k=1}^{n-1} a_k w_k \right\rangle \\
&= \sum_{1 \leqslant k \leqslant l \leqslant n-1} a_k a_l k(n-l)/n + \sum_{1 \leqslant l < k \leqslant n-1} a_k a_l l(n-k)/n \\
&\leqslant 2D^2.
\end{aligned}$$

Hence, by Theorem 2.4 in [7], the corresponding eigenvalue $-\lambda$ of Δ satisfies

$$\lambda = 2\langle \mathbf{v}, \sigma \rangle + \langle \sigma, \sigma \rangle \leqslant 2nD + 2D^2.$$

Proof. The proof proceeds by a similar computation to the proof of Lemma C.1. We use Theorem C.2 for SU(n) and pick a system of fundamental weights from [7]

□

C.3 Estimates of the eigenvalues in Sp(n)

For the bounds on the eigenvalues of the Laplace-Beltrami operator of Sp(n), we use the following system of fundamental weights (which are a scalar multiple of those in [12]). Setting $u_j = \mathbf{i}(E_{j,j} - E_{n+j,n+j})$ for each $j \in [n]$, we let

$$w_i = \sum_{j=1}^{i} u_i$$

be the ith fundamental weight, for each $i \in [n]$. The irreducible representation corresponding to $v = \sum_{i=1}^{n} r_i w_i$ corresponds to the Young diagram $\lambda = (\lambda_1, \ldots, \lambda_n)$, where $r_i = \lambda_i - \lambda_{i+1}$ for all $i \in [n]$ and $\lambda_{n+1} := 0$; the corresponding degree is

$$D = \sum_{i=1}^{n} \lambda_i = \sum_{j=1}^{n} j r_j. \tag{15}$$

From here on, the analysis is essentially the same as for SO_n (with n odd).

Using the formula in Theorem C.2, the eigenvalue corresponding to $v = \sum_{i=1}^{n} r_i w_i$ is

$$\lambda_v = -\|v + \rho\|_2^2 + \|\rho\|_2^2,$$

where here and throughout we think of the matrices v and ρ as $(2n)^2$ length vector by the natural flattening (equivalently $\langle A, B\rangle = \mathsf{tr}(A^t B)$). Then

$$\lambda_v = -2\langle v, \rho\rangle - \|v\|_2^2,$$

and we estimate the norm of v and the inner product between v and ρ. To compute the norm of v we write

$$v = \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} r_i \sum_{j=1}^{i} u_j = \sum_{j=1}^{n} \left(\sum_{i=j}^{n} r_i\right) u_j,$$

and since the u_j's are mutually orthogonal and each has 2-norm-squared equal to 2, we get that

$$\|v\|_2^2 = 2\sum_{j=1}^{n} \left(\sum_{i=j}^{n} r_i\right)^2 \leqslant 2\left(\sum_{j=1}^{n}\left(\sum_{i=j}^{n} r_i\right)\right)^2 = 2\left(\sum_{j=1}^{n}\sum_{i=1}^{n} i r_i\right)^2,$$

which is at most $2D^2$ by (15).

To estimate the inner product between v and ρ note that ρ is the vector v in which we take all r_i's to be 1, hence by the computation above

$$\rho = \sum_{i=1}^{n} w_i = \sum_{j=1}^{n} (n - j + 1)\, u_j,$$

and so

$$\langle \rho, v\rangle = 2\sum_{j=1}^{n} (n - j + 1)\left(\sum_{i=j}^{n} r_i\right) \leqslant 2n\sum_{j=1}^{n}\left(\sum_{i=j}^{n} r_i\right) \leqslant 2nD.$$

where in the last inequality we used (15). Overall, we get that the eigenvalue λ_v satisfies $\lambda_v \geqslant -2D^2 - nD$.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il