Inventiones mathematicae (2024) 237:1329–1375
https://doi.org/10.1007/s00222-024-01273-1

On the largest product-free subsets of the alternating groups

Peter Keevash[1] · Noam Lifshitz[2] · Dor Minzer[3]

Received: 16 February 2023 / Accepted: 17 May 2024 / Published online: 29 May 2024

Abstract
A subset A of a group G is called product-free if there is no solution to $a = bc$ with a, b, c all in A. It is easy to see that the largest product-free subset of the symmetric group S_n is obtained by taking the set of all odd permutations, i.e. $S_n \backslash A_n$, where A_n is the alternating group. In 1985 Babai and Sós (Eur. J. Comb. 6(2):101–114, 1985) conjectured that the group A_n also contains a product-free set of constant density. This conjecture was refuted by Gowers (whose result was subsequently improved by Eberhard), still leaving the long-standing problem of determining the largest product-free subset of A_n wide open. We solve this problem for large n, showing that the maximum size is achieved by the previously conjectured extremal examples, namely families of the form $\{\pi : \pi(x) \in I, \pi(I) \cap I = \varnothing\}$ and their inverses. Moreover, we show that the maximum size is only achieved by these extremal examples, and we have stability: any product-free subset of A_n of nearly maximum size is structurally close to an extremal example. Our proof uses a combination of tools from Combinatorics and Non-abelian Fourier Analysis, including a crucial new ingredient exploiting some recent theory developed by Filmus, Kindler, Lifshitz and Minzer for global hypercontractivity on the symmetric group.

1 Introduction

The problem of determining the largest product-free sets in groups was first raised by Babai and Sós [2] in 1985, who asked whether a finite group always contains a product-free set of constant density. They were particularly interested in the alternating group $A_n \subseteq S_n$, and they conjectured that A_n indeed contains such a large product-free set. The Babai-Sós conjecture was refuted by Gowers [8], who showed that any product-free set in A_n can have density at most $O(1/n^{1/3})$. Moreover, he demonstrated a more general phenomenon by showing that any *D-quasirandom*

✉ D. Minzer

1 Mathematical Institute, University of Oxford, Oxford, UK

2 Einstein Institute of Mathematics, Hebrew University, Jerusalem, Israel

3 Department of Mathematics, Massachusetts Institute of Technology, Cambridge, USA

group can only have product-free sets of density at most $O(1/D^{1/3})$. Here, a group is said to be D-quasirandom if the dimension of any non-trivial irreducible representation is at least D.

In the reverse direction, answering a question raised by Gowers, Nikolov and Pyber [12] showed that if a finite group G has a non-trivial irreducible representation of dimension D, then G does contain a product-free set of density $\Omega(1/D)$. Combined with Gowers' result one sees that the density of the largest product-free set in a group is determined by its quasirandomness, with some polynomial dependence that remains to be investigated. Thus, a key open problem is to determine the precise connection between the quasirandomness of a group and the density of its largest product-free sets.

Here, motivated by the original question of Babai and Sós [2], recently highlighted again by Green [9], we study the largest product-free sets in A_n. It was conjectured independently by Crane (personal communication) and Kedlaya [11] that the following sets (and their inverses) provide the extremal examples

$$F_I^x := \{\pi : \pi(x) \in I, \pi(I) \cap I = \varnothing\}.$$

Writing μ for uniform measure on A_n and $|I| = t\sqrt{n}$, one can calculate $\mu(F_I^x) \approx te^{-t^2}n^{-1/2}$. Improving earlier bounds of Kedlaya [11] and Gowers [8], the conjecture was proved up to logarithmic factors by Eberhard [4], who showed that any product-free $A \subseteq A_n$ has $\mu(A) = O(n^{-1/2}\log^{7/2} n)$. Besides being off by logarithmic factors, Eberhard's techniques do not seem to be sufficient to prove any structural results regarding the extremal families. Our main result here answers the question completely for large n, as follows.

Theorem 1.1 *Suppose n is sufficiently large and $A \subseteq A_n$ is a product-free subset of maximum size. Then A or A^{-1} is some F_I^x.*

In fact, our techniques allow us to prove structural results for any product-free set in A_n whose density is at least (say) n^{-100}, thus bypassing the "quasirandomness barrier" that has limited previous techniques in high rank groups [4, 8].

The notion of quasirandomness for groups appeared implicity in the work of Sarnak and Xue [15], and was later used by Bourgain and Gamburd [3] to show that most Cayley graphs over $SL_d(\mathbb{Z}/p\mathbb{Z})$ are expanders. It is a major open problem to prove analogues of the Bourgain-Gamburd theorem for finite simple groups of high rank. Indeed, their technique crucially relies on the fact that the quasirandomness of $SL_d(\mathbb{Z}/p\mathbb{Z})$ is polynomial in its size. Therefore, to handle finite simple groups of high rank, it is important to develop techniques that go beyond the quasirandomness parameter. The techniques presented herein do so for the group A_n: we combine ideas from harmonic analysis and representation theory to argue about sets whose density is $1/D^{\omega(1)}$.

1.1 Stability

Besides our exact result on the size of maximum product-free sets in A_n, we also obtain stability results describing the approximate structure of product-free sets that are somewhat large. The following ‘99% stability’ result shows that any moderately large product-free subset of A_n is essentially contained in an extremal family; note that it applies to sets whose measure is much smaller than the extremal family.

Theorem 1.2 *Suppose n is sufficiently large and $A \subseteq A_n$ is a product-free set with $\mu(A) \geq n^{-0.66}$. Then there is some F_I^x such that $\mu(A \backslash F_I^x) < n^{-0.66}$ or $\mu(A^{-1} \backslash F_I^x) < n^{-0.66}$.*

Moreover, if $A \subseteq A_n$ is a product-free set with size very close to the maximum we show that A or A^{-1} is contained in some F_I^x.

Theorem 1.3 *There exists an absolute constant c such that if n is sufficiently large and $A \subseteq A_n$ is a product-free set with $\mu(A) > \max_{I,x} \mu\left(F_I^x\right) - \frac{c}{n}$ then there is some F_I^x such that $A \subseteq F_I^x$ or $A^{-1} \subseteq F_I^x$.*

We also study the following ‘1% stability’ problem.

Problem 1.4 Suppose that $A \subseteq A_n$ is a product-free set of density $> 1/n^C$ for an absolute constant C. What can be said about the structure of A?

Dictators and *t*-umvirates. Here we describe the structures appearing in the answer to this problem. A set of the form $\mathcal{D}_{i \to j} = \{\sigma \in A_n : \sigma(i) = j\}$ is called a *dictator*. Let $D_1, \ldots D_t$ be distinct dictators that have a nonempty intersection. Following Friedgut [7], we call their intersection a *t-umvirate*. Equivalently, a t-umvirate $\mathcal{U}_{I \to J}$ corresponding to ordered sets $I = (i_1, \ldots, i_t)$, $J = (j_1, \ldots, j_t)$ is the set of permutations that send each i_k to j_k.

Densities. Let $A, B \subseteq S_n$. The *density* of A inside B is $\mu_B(A) := \frac{|A \cap B|}{|B|}$. To restrict $A \subseteq S_n$ to $\mathcal{U}_{I \to J}$ we write $A_{I \to J} = A \cap \mathcal{U}_{I \to J}$ and $\mu(A_{I \to J}) = \frac{|A \cap \mathcal{U}_{I \to J}|}{|\mathcal{U}_{I \to J}|}$ for its density in the ambient space $\mathcal{U}_{I \to J}$.

Product-free sets correlate with *t*-umvirates. Our next ‘1% stability’ theorem shows that any product-free set that is somewhat dense has some local structure. This is analogous to the strong local structure exhibited by the extremal families F_I^x, which have $\Theta(1)$ density inside each dictator $1_{x \to i}$ with $i \in I$, as when $|I| = \Theta(\sqrt{n})$ a random permutation sends I to its complement with probability $\Theta(1)$. We show that a similar, albeit weaker, phenomenon holds for product-free sets with any polynomial density that can be much smaller than that in the extremal examples: such sets have a density bump inside a t-umvirate.

Theorem 1.5 *Fix $r \in \mathbb{N}$, suppose n is sufficiently large, and $A \subseteq A_n$ is product-free with $\mu(A) > n^{-r}$. Then there exists some t-umvirate with $t \leq 4r$ in which A has density at least $n^{t/4}\mu(A)$.*

The proof will use the trace method, a recent level-d inequality due to Filmus, Kindler, Lifshitz and Minzer [6], and novel upper bounds on eigenvalues of Cayley graphs over the symmetric group.

Globalness. Theorem 1.5 naturally leads us to define the notion of 'globalness', which plays a crucial role throughout the paper. This is intuitively a pseudorandomness property stating that membership is not determined by local information, which one can think of as being the polar extreme to dictators.

More precisely, we make the following definition, saying that small restrictions do not have large measure. We include two versions of the same concept with different parameterisations, as we need the first version when referring to [6], but the second version is more natural for the applications in our paper.

Definition 1.6 We say $A \subseteq A_n$ is (t, ϵ)-global if the density of A inside each t-umvirate is $< \epsilon^2$. We say that A is relatively (t, K)-global if the density of A inside each t-umvirate is at most $K\mu(A)$.

1.2 Product-free triples

We also consider the following cross version of the question. Given $A, B, C \subseteq A_n$, we say that the triple (A, B, C) is product-free if there is no solution of[1] $ab = c$ with $a \in A$, $b \in B$, $c \in C$. In particular, (A, A, A) is product-free if and only if A is product-free. We consider the following problem.

Problem 1.7 What are the possible sizes of product-free triples?

Problem 1.7 incorporates the well-studied problem of upper bounding sizes of independent sets inside Cayley graphs (see e.g. Ellis, Friedgut and Pilpel [5]), which concerns the special case $A = A^{-1}$ and $B = C$.

For simplicity we restrict ourselves to the question of maximising $\min(\mu(A), \mu(B), \mu(C))$ when (A, B, C) is product-free. Gowers [8] established an upper bound of $(n-1)^{-1/3}$ by showing that if (A, B, C) is product-free then $\mu(A)\mu(B)\mu(C) \le \frac{1}{n-1}$. Eberhard [4] improved this to $O(n^{-1/2}\log^{3.5} n)$ by showing that one of $\mu(A)\mu(B)$, $\mu(A)\mu(C)$, $\mu(B)\mu(C)$ is $O(n^{-1}\log^7 n)$. We obtain the following bound, which gives the correct order of magnitude, as shown by examples of the form

$$A = 1_{I\to\overline{I}} := \{\pi : \pi(I) \cap I = \varnothing\} \qquad C = B = 1_{x\to I} := \{\pi : \pi(x) \in I\}.$$

Theorem 1.8 *Let n be sufficiently large and $A, B, C \subseteq A_n$ with (A, B, C) product-free. Then one of $\mu(A)$, $\mu(B)$, $\mu(C)$ is at most $100\sqrt{\log n}/\sqrt{n}$.*

In fact, we show corresponding stability results that are a bit harder to state (see Theorem 6.4). Roughly speaking, they say that any sufficiently dense product-free triple has the approximate form $\left(1_{I\to\overline{J}}, 1_{x\to I}, 1_{x\to J}\right)$. Moreover, we also prove a version of Theorem 1.5 that finds structure in product-free triples that are only polynomially dense.

As discussed later in the introduction, our approach can be viewed as an non-abelian analogue of Roth's bound for sets of integers with no three-term arithmetic

[1] To clarify our notation for composition of permutations: we mean that $a(b(x)) = c(x)$ for all $x \in [n]$.

progression, whereby we improve the earlier approaches of Gowers and Eberhard by establishing a form of the 'Structure versus Randomness' dichotomy. We achieve this by exploiting some recent theory developed by Filmus, Kindler, Lifshitz and Minzer [6] for global hypercontractivity on the symmetric group, and by also developing some further theory of the 'Cayley operators' associated to global subsets.

1.3 Techniques

We call a set of the form $1_{x\to I} = \{\sigma : \sigma(x) \in I\}$ a *star* and $\left\{\sigma : \sigma^{-1}(x) \in I\right\} = 1_{x\to I}^{-1}$ an *inverse star*.

The main steps in the proof of our main theorem (Theorem 1.1) are as follows.

1. Dictatorial structure: We show that A has large density inside many dictators. In fact, we show that in some sense, the product freeness of A is completely explained by its densities inside dictators.
2. Star structure: We upgrade our dictatorial structure into a tighter star structure, by finding some S that is either a star or an inverse star such that $|A\setminus S|$ is small and A has significant density in each restriction defined by S.
3. Bootstrapping: Using the approximate star structure, we deduce our exact results from further stability analysis showing that any small deviation from the structure leads to a suboptimal configuration.

Gowers' approach: the second eigenvalue. We start out along the path established by Gowers [8]. His idea was to express the number of products in $A \subseteq A_n$ of density α as the sum of a 'main term' and an 'error term', where the error term is smaller in magnitude than the main term when α is large, so the number of products cannot be zero. The main term $\alpha^3|A_n|^2$ is the expected number of products in a random set of density α, whereas his bound for the error term is $(n-1)^{-1/2}\alpha^{3/2}|A_n|^2$, which for product-free A implies $\alpha \le (n-1)^{-1/3}$.

We will now outline his argument. Let $f = 1_A$ be the indicator function of $A \subseteq A_n$ and T be the linear operator on $L^2(A_n)$ defined by

$$(Tg)(\pi) = \mathbb{E}_{\sigma\sim A_n}\left[f(\sigma)\, g(\sigma\circ\pi)\right].$$

Then A contains a product if and only if $\langle f, Tf\rangle > 0$. Let V' be the space of functions of expectation 0. Then we have $\langle f, Tf\rangle = \alpha^3 + \left\langle f', Tf'\right\rangle$, where $f' = f - \alpha \in V'$. Writing T^* for the adjoint of T, some Representation Theory of S_n (discussed in more detail below) tells us that the self-adjoint operator T^*T acts on V' with all eigenvalues bounded absolutely by $\frac{\alpha}{n-1}$. We deduce

$$\begin{aligned}\left\langle f', Tf'\right\rangle^2 &\le \|f'\|_2^2\|Tf'\|_2^2 = \|f'\|_2^2\left\langle f', T^*Tf'\right\rangle\\ &\le \|f'\|_2^3\|T^*Tf'\|_2 \le \|f'\|_2^4\alpha/(n-1) \le \alpha^3/(n-1).\end{aligned}$$

Thus we can deduce $\langle f, Tf\rangle > 0$ if $\alpha^3 > \sqrt{\alpha^3/(n-1)}$, i.e. if $\alpha > (n-1)^{-1/3}$.

Beyond spectral gap: the degree decomposition. This path was continued by Eberhard [4], whose improved bound is based on a refined analysis of the main contribution to the error term (he credits Ellis and Green for suggesting this approach), replacing the basic decomposition $f = \alpha + f'$ by a refined 'level' decomposition $f = \sum_{d=0}^{n-1} f^{=d}$.

Instead of working with functions on A_n, henceforth it will be more convenient to work with functions on S_n that are supported on A_n, so we will now let $\alpha = |A|/n!$ denote the density of A inside S_n. This allows us to import machinery developed for S_n without reworking it for A_n, although it introduces some slight inconvenience in keeping track of factors of 2 in the calculations.

For $d \le n$, let W_d be the linear subspace of $L^2(S_n)$ generated by indicators of d-umvirates. The degree of a nonzero function f is the minimal d such that $f \in W_d$. We define $V_{=d} = W_d \cap W_{d-1}^{\perp}$ and note that the spaces $V_{=d}$ form an orthogonal decomposition of $L^2(S_n)$, known as the *degree* decomposition of S_n. For a function f on S_n we write $f^{=d}$ for the projection of f onto $V_{=d}$.

To set up Eberhard's refined analysis, we write $f = 2\alpha + 2f^{=1} + f''$ with

$$f'' \in V'' = \left\{ g - 2\mathbb{E}[g] - 2g^{=1} : \text{g is supported on } A_n \right\}.$$

Again, some Representation Theory shows that T^*T acts on V'' with all eigenvalues $O(\alpha/n^2)$. Similarly to the above argument this implies

$$\langle f, Tf \rangle = 2\alpha^3 + 2\left\langle f^{=1}, Tf^{=1} \right\rangle + \left\langle f'', Tf'' \right\rangle,$$

where $\left\langle f'', Tf'' \right\rangle = O(\alpha^{3/2}/n)$ is negligible compared to α^3 when $\alpha \gg n^{-2/3}$. Thus it suffices to control the 'linear term' $\left\langle f^{=1}, Tf^{=1} \right\rangle$, which turns out to be equal to

$$\mathbb{E}_{\sigma,\pi \sim S_n} \left[f^{=1}(\sigma) f^{=1}(\pi) f^{=1}(\sigma\pi) \right].$$

However, it is not generally true that the linear term is small compared with the main term. Indeed, this would imply 'product mixing', i.e. that the number of products is close to that in a random set of the same density, but Eberhard [4] constructed examples with significantly more products when $\alpha = o\left(n^{-1/3}\right)$. The key to his approach is that the above counting argument only needs a lower bound on the linear term, and that this exhibits much nicer concentration properties than the upper bound. Nevertheless, even these better estimates break down for densities within a polylogarithmic factor of the optimum bound.

1.4 Our approach: structure versus randomness

Our approach to stability can be viewed as a further refinement of this method that is close in spirit to Roth's theorem, in that it is analogous to the 'structure versus randomness' dichotomy: for Roth's theorem, if the error term is large then A correlates with an arithmetic progression, whereas in our setting, if the linear term cancels the main term then A correlates with dictators.

Our starting point for this strategy is the formula

$$f^{=1}(\pi) = \sum_{i,j \in [n]} a_{ij} x_{i \to j},$$

where $x_{i\to j}(\pi) = 1_{\pi(i)=j}$, and each

$$a_{ij} = (1 - 1/n)(\mu(A_{i\to j}) - \mu(A))$$

measures the correlation of f with a dictator. A large value of $a_{i,j}$ corresponds to a large density inside a dictator. On the other hand, having all $a_{i,j}$ of the same order of magnitude as $\mu(A)$ can be interpreted as pseudorandomness. Some calculations reveal that

$$\|f^{=1}\|_2^2 = \tfrac{1}{n-1} \sum_{i,j\in[n]} a_{ij}^2,$$

and that

$$\mathbb{E}_{\sigma,\pi\sim S_n}\left[f^{=1}(\sigma)\, f^{=1}(\pi)\, f^{=1}(\sigma\pi)\right] = \frac{1}{(n-1)^2} \sum_{i,j,k\in[n]} a_{ij}a_{jk}a_{ik}.$$

These formulae alone do not suggest any structural properties for a contribution of $-\alpha^3$ to the left hand side; in principle, $\Theta(n/\alpha)$ values of $a_{i,j} = -\alpha$ could contribute $\frac{1}{n^2}\Theta(n/\alpha)^{3/2}(-\alpha)^3$, so when $\alpha = o\left(n^{-1/3}\right)$ we seem to have no structure.

Such extreme situations can be ruled out by concentration of measure, which is a key tool in Eberhard's approach, but seems doomed to give up logarithmic factors. However, much stronger structure can be extracted from a recent hypercontractive inequality of Filmus, Kindler, Lifshitz, and Minzer [6].

It will be convenient for us to put the coefficients a_{ij} inside a matrix $\mathrm{A} = \left(a_{ij}\right)$ and to equip real valued $n \times n$ matrices with the inner product $\left\langle \left(m_{ij}\right), \left(n_{ij}\right)\right\rangle = \sum_{i=1}^n \sum_{j=1}^n m_{ij}n_{ij}$. We then have

$$\|f^{=1}\|_2^2 = \frac{1}{n-1}\|\mathrm{A}\|_2^2$$

and

$$\mathbb{E}_{\sigma,\pi}\left[f^{=1}(\pi)\, f^{=1}(\sigma)\, f^{=1}(\sigma\pi)\right] = \frac{1}{(n-1)^2}\left\langle \mathrm{A}^2, \mathrm{A}\right\rangle.$$

Our idea is to decompose $\mathrm{A} = \mathrm{A}_{\text{rand}} + \mathrm{A}_{\text{struc}} + \mathrm{A}_{-}$ as follows. We let the matrix A_{-} consist of the negative coefficients a_{ij}, where the other coefficients are replaced by 0. We then set the matrix $\mathrm{A}_{\text{struc}}$ to consist of the 'large' values $a_{ij} \geq \epsilon$, for some carefully chosen $\epsilon > 0$. Finally, we let A_{rand} consist of the 'small' positive coefficients $a_{ij} \in (0, \epsilon)$.

We then expose the dictatorial structure of A by showing that

$$\mathbb{E}_{\sigma,\pi}\left[f^{=1}(\pi)\, f^{=1}(\sigma)\, f^{=1}(\sigma\pi)\right] \geq \langle \mathrm{A}_{\text{struc}}\mathrm{A}_{-}, \mathrm{A}_{\text{struc}}\rangle + \langle \mathrm{A}_{-}\mathrm{A}_{\text{struc}}, \mathrm{A}_{\text{struc}}\rangle + \left\langle \mathrm{A}_{\text{struc}}^2, \mathrm{A}_{-}\right\rangle + o\left(\alpha^3\right). \quad (1)$$

This could be understood intuitively as follows. Writing $AB = \{\sigma\tau : \sigma \in A, \tau \in B\}$, we see that the dictators satisfy $\mathcal{D}_{j\to k} \cdot \mathcal{D}_{i\to j} = \mathcal{D}_{i\to k}$. When expanding (1) in terms of the coefficients a_{ij}, it shows that the only significant negative contribution to

$$\mathbb{E}_{\sigma,\pi}\left[f^{=1}(\pi)\, f^{=1}(\sigma)\, f^{=1}(\sigma\pi)\right]$$

comes from triples $a_{jk}a_{ij}a_{ik}$ corresponding to dictators $\mathcal{D}_{j\to k} \cdot \mathcal{D}_{i\to j} = \mathcal{D}_{i\to k}$, with the property that A has a large density in two of the dictators and a small density in the remaining one.

To establish (1) we will expand the inner product $\left\langle \mathrm{A}^2, \mathrm{A}\right\rangle$ in terms of the matrices A_-, $\mathrm{A}_{\mathrm{struc}}$, $\mathrm{A}_{\mathrm{rand}}$ and use the inequality

$$\langle MN, S\rangle \le \|M\|_2\|N\|_2\|S\|_2$$

to upper bound the undesirable terms. Our proof will thus crucially rely on upper bounds for $\|\mathrm{A}_{\mathrm{rand}}\|_2$ and $\|\mathrm{A}_-\|_2$, which we will establish via a hypercontractive inequality for global functions, as discussed in the next subsection.

Level-1 inequalities. The relationship between hypercontractive inequalities and inequalities that upper bound the ‘level-1 weight’ $\|f^{=1}\|_2^2$ is well-known in the context of Boolean functions $f\colon \{0,1\}^n \to \{0,1\}$. The inequality

$$\|f^{=1}\|_2^2 \le 2\mathbb{E}^2[f]\log\left(\frac{1}{\mathbb{E}[f]}\right)$$

is true for all Boolean functions and is known as the *level*-1 *inequality* (see O’Donnell [13, Chap. 5]). As

$$\frac{1}{(n-1)^2}\left\langle \mathrm{A}^2, \mathrm{A}\right\rangle \le \frac{\|\mathrm{A}\|_2^3}{(n-1)^2} = \frac{\|f^{=1}\|_2^3}{\sqrt{n-1}},$$

a similar level-1 inequality for the symmetric group would be most desirable, as it would imply that $\frac{\|f^{=1}\|_2^3}{\sqrt{n-1}}$ is negligible compared to α^3. However, such an inequality is not true in general, as it fails for dictators, and more generally for t-umvirates with small t.

The local nature of the obstructions suggests that the approach could be rescued by proving a level-1 inequality for global functions. Indeed, this was achieved in the analogous setting of general product spaces by Keevash, Lifshitz, Long and Minzer [10], who developed a hypercontractivity theory for global functions that has recently been a fruitful source of many applications. The corresponding results in the setting of the symmetric group have recently been established by Filmus, Kindler, Lifshitz and Minzer [6]. Their level-1 inequality shows that if $f\colon S_n \to \{0,1\}$ is $(2,\epsilon)$-global then

$$\|f^{=1}\|_2^2 \le 2\epsilon\mathbb{E}[f]\log^{O(1)}\left(\frac{1}{\epsilon}\right).$$

This inequality cannot be applied directly to our setting, as we cannot guarantee that f has small density inside each duumvirate. However, we are able to extend their approach to obtain the upper bounds

$$\|A_-\|_2 \le \mathbb{E}[f] \log^{O(1)}(2/\mathbb{E}[f])$$

and

$$\|A_{\text{rand}}\|_2^2 \le \epsilon \mathbb{E}[f] \log^{O(1)}\left(\frac{2}{\epsilon}\right).$$

The key idea is to apply the hypercontractive result of [6] to

$$f_- := \sum_{a_{ij}<0} a_{ij}\left(x_{i\to j} - \frac{1}{n}\right)$$

and to

$$f_{\text{rand}} := \sum_{0<a_{ij}<\epsilon} a_{ij}\left(x_{i\to j} - \frac{1}{n}\right).$$

These inequalities will establish (1). As an expository simple case of our argument, in the next section we will show that this can be used to recover Eberhard's result, by setting $\epsilon = 1$, so that $A_{\text{struc}} = 0$. Extracting the star structure for smaller ϵ is considerably more complicated, so we defer an overview of this part of the argument to Sect. 5.3.

1.5 From star structure to extremal families

Once we know that A is almost contained in a star or inverse star S, say $S = 1_{x\to I}$, then it is not too hard to show that it is in fact almost contained in F_I^x. In other words, we wish to show for each $i, i' \in I$ that A has small density inside the dictator $\mathcal{D}_{i\to i'}$. We accomplish this by inspecting the triple $(A_{i\to i'}, A_{x\to i}, A_{x\to i'})$. Such triples should be intuitively considered as product-free after factoring out the corresponding dictators with

$$\mathcal{D}_{i\to i'}\mathcal{D}_{x\to i} = \mathcal{D}_{x\to i'}.$$

To formalise this, we would like a variant of Eberhard's bound that holds not only for product-free triples $A', B', C' \subseteq A_n$, but also for product-free sets A', B', C' living inside compatible dictators. This can be achieved by the following transformation. We set $A' = \left(i'\,n\right) A\,(n\,i)$, $B' = (i\,n)\,A\,(n\,x)$, and $C' = \left(i'\,n\right) A(x\,n)$. The transformation from (A, B, C) to $\left(A', B', C'\right)$ preserves products. Moreover, the restrictions

$$\left(A'_{n\to n}, B'_{n\to n}, C'_{n\to n}\right)$$

correspond to the original restrictions

$$(A_{i\to i'}, B_{x\to i}, C_{x\to i'}).$$

We may then view the triple $\left(A'_{n\to n}, B'_{n\to n}, C'_{n\to n}\right)$ as subsets of A_{n-1} and translate Eberhard's bounds inside A_{n-1} to the densities of $A_{i\to i'}$, $B_{x\to i}$ and $C_{x\to i'}$.

Similar considerations combined with some more involved structural arguments will show that $\mu\left(A \setminus F_I^x\right)$ is much smaller than $\mu\left(F_I^x \setminus A\right)$, thereby showing that F_I^x is extremal.

1.6 Eigenvalues of global Cayley graphs

The main tool for proving our 1% stability result (Theorem 1.5) is an upper bound on the eigenvalues of Cayley graphs $\mathrm{Cay}(S_n, A)$ that correspond to global sets A. We believe that the result has independent interest and will be beneficial in various other applications; indeed, the study of Cayley graphs on S_n and their eigenvalues is the basis of an entire field of research (see e.g. Diaconis [14]).

Let $A = A^{-1}$ and let T_A be the operator corresponding to the random walk on $\mathrm{Cay}(G, A)$, given by $T_A f(x) := \mathbb{E}_{a\sim A}[f(ax)]$. As mentioned above, the operator T_A preserves each of the spaces $V_{=d}$. We define

$$r_d(T_A) = \sup_{f\in V_{=d}} \frac{\|\mathrm{T}_A f\|_2}{\|f\|_2},$$

or equivalently $r_d\left(T_A\right)^2$ is the largest eigenvalue of the self-adjoint operator $T_A^* T_A$ acting on $V_{=d}$.

A useful fact from representation theory gives lower bounds on the dimension of each eigenspace of T_A using the observation that each is invariant under the action of S_n from the right. Indeed, this was a crucial ingredient in the seminal work of Ellis, Friedgut and Pilpel [5], who derived bounds on the dimensions from well-known properties of invariant spaces; in particular, the dimensions are $\Theta(n^d)$ when $d = O(1)$.

The trace method provides a fundamental way to exploit this lower bound. Specifically, writing m_λ for the multiplicity of an eigenvalue λ of T_A, we have

$$\frac{1}{\mu(A)} = \mathrm{tr}(T_A^* T_A) = \sum_{\lambda\in\mathrm{spec}(T_A)} m_\lambda \lambda^2.$$

When A is dense (i.e. $\mu(A) = \Theta(1)$) we thus obtain $|\lambda| = O(1/\sqrt{m_\lambda})$, and so $r_d(T_A) = O_d(1/n^{d/2})$. Furthermore, if A is closed under conjugation then the operator T_A commutes with the action of S_n from both sides, in which case we have $m_\lambda \geq \Omega(n^{2d})$, giving the stronger bound $r_d(T_A) = O_d(1/n^d)$. We show that similar statements hold for global sets, even when they are quite sparse.

Theorem 1.9 *Let $C, d > 0$ and suppose that n is sufficiently large. Suppose that $A \subseteq S_n$ is relatively $(2d, K)$-global with $\mu\left(A\right) \geq n^{-C}$. Then*

$$r_d(T_A) \leq \frac{\sqrt{K}\log^{O(d)} n}{n^{d/2}}.$$

Furthermore, if A is also closed under conjugation then

$$r_d(T_A) \leq \frac{\sqrt{K}\log^{O(d)}(n)}{n^d}.$$

Theorem 1.9 is the crucial new ingredient for proving our 1% stability result for product-free sets of density $\geq n^{-C}$ (Theorem 1.5).

1.7 Organisation of the paper

We start in the next section with some technical preliminaries on the representation theory of S_n and also the proofs of the results discussed immediately above, i.e. new eigenvalue estimates for global sets and their application to the proof of our 1% stability result. Section 3 considers the analysis of linear functions on the symmetric group. The main result of this section will be a level-1 inequality for the pseudorandom part of a function, which in itself will suffice to recover Eberhard's result. We then move into more refined arguments that extract structural properties of product-free sets, first exposing the dictatorial structure in Sect. 4 and then the precise star structure in Sect. 5. In Sect. 6 we implement our bootstrapping arguments that refine the approximate structure to deduce our main results, giving the exact extremal result and strong stability results for product-free sets in A_n. The final section contains some concluding remarks.

2 1% stability

This section contains some background on the representation theory of S_n, our new eigenvalue estimates for global sets and their application to the proof of our 1% stability result.

2.1 Notation

We write $X = O(Y)$ if there exists an absolute constant $C > 0$ such that $X \leq C \cdot Y$, and similarly $X = \Omega(Y)$ if there exists an absolute constant $c > 0$ such that $X \geq c \cdot Y$. We write $X = \Theta(Y)$ if$X = O(Y)$ and $X = \Omega(Y)$. We also write $X \leq Y \log^{O(1)} n$ if there exists an absolute constant C such that $X \leq Y \log^C n$.

For $A, B \subseteq S_n$ we write

$$\mu_A(B) = \frac{|A \cap B|}{|A|} = \Pr_{a \sim A}[a \in B].$$

Here $a \sim A$ means that a is uniformly distributed in A. We also write $\sigma, \tau \sim S_n$ to mean that σ, τ are independent and uniformly distributed in S_n.

We discuss the space of real-valued functions on S_n equipped with expectation inner product and L^p-norms. For a function $f\colon S_n \to \mathbb{R}$ we often write $\mathbb{E}[f]$ to denote $\mathbb{E}_{\sigma \sim S_n}[f(\sigma)]$, although we caution the reader that this usage will depend on context, as when we define $\mathbb{E}[f_{I \to J}]$ below the expectation will be conditioned on the given restriction.

2.2 Restrictions

We define restrictions of functions in a manner that naturally generalizes the notion of restrictions of subsets of S_n used in the introduction.

Definition 2.1 Let $I = (i_1, \ldots, i_t)$, $J = (j_1, \ldots, j_t) \subseteq [n]$ be ordered sets of size t. We denote by $\mathcal{U}_{I \to J}$ the t-umvirate of permutations sending each i_l to j_l.

We define $x_{i \to j} : S_n \to \{0, 1\}$ by

$$x_{i \to j}(\pi) = 1_{\pi(i)=j}.$$

Let $I \subseteq [n]$, $J \subseteq [n]$ be ordered sets of some size t. We denote by $f_{I \to J} : \mathcal{U}_{I \to J} \to \mathbb{R}$ the restriction of f to $\mathcal{U}_{I \to J}$. We write

$$\mathbb{E}[f_{I \to J}] = \mathbb{E}_{\sigma \sim S_n}[f(\sigma) | \sigma \in \mathcal{U}_{I \to J}]$$

and

$$\| f_{I \to J} \|_2^2 = \mathbb{E}_{\sigma \sim S_n}\left[f^2(\sigma) | \sigma \in \mathcal{U}_{I \to J}\right].$$

We may identify $\mathcal{U}_{I \to J}$ with S_{n-t} via any fixed permutations σ, π such that $\sigma(n - t + l) = i_l$ and $\pi(j_l) = n - t + 1$ for each $l \in [t]$. Then $\pi \mathcal{U}_{I \to J} \sigma$ is the set of permutations on S_n fixing $n - t + l$ for each $l \in [t]$, which can be identified with S_{n-t}. We will use this identification to import results on the symmetric group S_{n-t} to the t-umvirate $\mathcal{U}_{I \to J}$.

2.3 Orthogonal decompositions

Our proof will use spectral analysis over S_n, so we need to recall its level decomposition and the more refined representation theoretical decomposition into isotypical components.

The level decomposition. We start by decomposing according to degree, as discussed in the introduction.

Definition 2.2 The space W_d is the linear span of the indicators of the d-umvirates. We say that a real-valued function on S_n has *degree* at most d if it belongs to W_d.

By construction, $W_{d-1} \subseteq W_d$ for all $d \geq 1$. We define the space of functions of pure degree d as $V_{=d} = W_d \cap W_{d-1}^{\perp}$. It is easy to see that $V_n = V_{n-1}$, and so we can decompose each real-valued function $f : S_n \to \mathbb{R}$ as $f = f^{=0} + f^{=1} + \cdots + f^{=n-1}$, where $f^{=i} \in V_{=i}$. We refer to this decomposition as the *level decomposition* of f.

Sometimes we require the following finer decomposition that decomposes each $f^{=i}$ into more structured pieces.

The representation theoretic decomposition. Here we will list the properties we require of the finer decomposition of functions on S_n into isotypical components; these can be found e.g. in [6, Sect. 7.2].

We adopt the following standard notation. We write $\lambda \vdash n$ to denote that λ is a partition of n. A partition λ uniquely corresponds to a Young diagram. Its transpose λ^t

is obtained by swapping the roles of the rows and the columns of the Young diagram. We let S_n act on $L^2(S_n)$ from the left by $(\tau, f) \mapsto {}^\tau f$ where ${}^\tau f(\sigma) = f(\tau^{-1}\sigma)$ and from the right by $(f, \tau) \mapsto f^\tau$ where $f^\tau(\sigma) = f(\sigma\tau^{-1})$.

Lemma 2.3 *There exists an orthogonal decomposition $L^2(S_n) = \bigoplus_{\lambda\vdash n} V_{=\lambda}$ with the following properties*:

1. *For some numbers* $\dim(\lambda)$, *each $V_{=\lambda}$ is the direct sum of* $\dim(\lambda)$ *irreducible representations of dimension* $\dim(\lambda)$.
2. *If* $\mathrm{T}\colon L^2(S_n) \to L^2(S_n)$ *commutes with the action of S_n either from the left or from the right then $\mathrm{T}V_{=\lambda} \subseteq V_{=\lambda}$, and so $\mathrm{T}V_{=d} \subseteq V_{=d}$.* (*See the proof of* [6, *Claim* 7.4].)
3. *If* T *is self adjoint and commutes with the action from one side then the dimension of each eigenspace of* T *inside $V_{=\lambda}$ is at least* $\dim(\lambda)$. *If it commutes with the action of S_n from both sides then $V_{=\lambda}$ is contained in an eigenspace of* T. (*Both follow from Schur's Lemma. See* [6, *Claim* 7.5] *for the first. The second uses irreducibility as an $S_n \times S_n$ module of the isotypic component $End(V_{=\lambda})$*.)
4. *$V_{=\lambda} \le V_{=d}$ if and only if the first row of λ is of length $n - d$. In this case, we write $d_\lambda = d$ and $\tilde{d}_\lambda = \min(d_\lambda, d_{\lambda^t})$.*
5. *Multiplication by the sign character sends V_λ to V_{λ^t}.* (*See* [6, *Lemma* 7.3].)
6. *If n is sufficiently large, $d < n/10$ and $\tilde{d}_\lambda > d$ then* $\dim(\lambda) > \left(\frac{n}{ed}\right)^d$. (*See* [6, *Lemma* 7.7].)

We write $f^{=\lambda}$ for the projection of f onto V_λ. We identify functions $f\colon A_n \to \{0, 1\}$ with function on S_n whose value is 0 on the odd permutations. Such functions satisfy $\mathsf{sign} \cdot f = f$. This gives rise to two decompositions of f as a sum of elements in V_λ, which therefore must be equal. The first is $f = \sum_{\lambda\vdash n} f^{=\lambda}$ and the second is $\sum_{\lambda\vdash n} \mathsf{sign} \cdot f^{=\lambda^t}$. This shows that $\mathsf{sign} \cdot f^{=\lambda} = f^{=\lambda^t}$.

2.4 Operators from functions

We write $\mathbb{E}[f]$ for $\mathbb{E}_{\pi\sim S_n}[f]$. Thus if $f = 1_A$ for $A \subseteq A_n$ then $\mathbb{E}[f] = |A|/n!$ denotes the density of A in S_n, not A_n. For $f \in L^2(S_n)$ we define operators L_f and R_f on $L^2(S_n)$ by

$$\mathrm{L}_f g(\sigma) = \mathbb{E}_\pi[f(\pi) g(\pi\sigma)] \quad \text{and} \quad \mathrm{R}_f g(\sigma) = \mathbb{E}_\pi[g(\sigma\pi) f(\pi)].$$

When $f = \frac{1_A}{\mu(A)}$, the operator L_f is the operator corresponding to the random walk sending σ to $a\sigma$ for a random $a \in A$. Similarly, R_f corresponds to the random walk that sends σ to σa. The operators L_f and R_f commute with the actions of S_n from one side. Indeed, for any $g \in L^2(S_n)$ and $\sigma, \tau \in S_n$ we have

$$\left(\mathrm{L}_f(g)\tau\right)(\sigma) = \mathrm{L}_f g\left(\sigma\tau^{-1}\right) = \mathbb{E}_\pi\left[f(\pi) g(\pi\sigma\tau^{-1})\right] = \mathrm{L}_f(g\tau)(\sigma).$$

Similarly, we have $\mathrm{R}_h(\tau g) = \tau(\mathrm{R}_h g)$. When f is a class function (meaning that $f(\sigma)$ depends only on the conjugacy class of σ) we have $\mathrm{L}_f = \mathrm{R}_f$. Indeed, when f

is a class function we have

$$\mathrm{L}_f g(\sigma) = \mathbb{E}_\pi\left[f(\pi)\, g(\pi\sigma)\right] = \mathbb{E}_\pi\left[f\left(\sigma^{-1}\pi\sigma\right) g(\pi\sigma)\right] = \mathbb{E}_\pi\left[f(\pi)\, g(\sigma\pi)\right]$$
$$= R_f g(\sigma).$$

Thus for a class function $\mathrm{L}_f = \mathrm{R}_f$ commutes with the S_n actions on both sides.

As mentioned, the trace method plays a crucial role in our work, and computing the trace of the operator $\mathrm{L}_f^*\mathrm{L}_f$ will allow us to upper bound its eigenvalues.

Lemma 2.4 *The operators* $\mathrm{L}_f^*\mathrm{L}_f$ *and* $\mathrm{R}_f^*\mathrm{R}_f$ *both have trace* $\|f\|_2^2$.

Proof We only consider $\mathrm{T} = \mathrm{L}_f$, as the proof for R_f is similar. We have

$$\begin{aligned}\operatorname{tr}\left(\mathrm{T}^*\mathrm{T}\right) &= \sum_{\pi\in S_n} n!\left\langle \mathrm{T}^*\mathrm{T}1_\pi, 1_\pi\right\rangle \\ &= \sum_{\pi\in S_n} n!\left\langle \mathrm{T}1_\pi, \mathrm{T}1_\pi\right\rangle = \sum_{\pi\in S_n}\sum_{\sigma\in S_n} \mathrm{T}1_\pi(\sigma)^2 \\ &= \sum_{\pi\in S_n}\sum_{\sigma\in S_n}\left(\mathbb{E}_{\tau\sim S_n}\left[f(\tau)\, 1_\pi(\tau\sigma)\right]\right)^2 \\ &= \left(\frac{1}{n!}\right)^2 \sum_{\pi\in S_n}\sum_{\sigma\in S_n} f\left(\pi\sigma^{-1}\right)^2 = \|f\|_2^2. \end{aligned}$$

□

2.5 Hypercontractivity of global functions

In this subsection we state two inequalities from [6]. To do so, we need the following natural extension of the definition of globalness from sets to functions.

Definition 2.5 We say $f: S_n \to \mathbb{R}$ is (d, ϵ)-global if for every ordered $I, J \subseteq [n]$ of size d we have $\|f_{I\to J}\|_2 \le \epsilon$. We say f is relatively (d, K)-global if $\|f_{I\to J}\|_2^2 \le K\|f\|_2^2$ for each such I, J.

The hypercontractivity inequality of [6] takes the following form.

Theorem 2.6 *There exists an absolute constant* C *such that the following holds. Let* $d, q \in \mathbb{N}$ *with* $q \ge 2$ *and* $n \ge q^{C\cdot d^2}$. *If* $f: S_n \to \mathbb{R}$ *is* $(2d, \epsilon)$*-global then*

$$\|f\|_q \le q^{O(d^2)} \epsilon^{1-\frac{2}{q}} \|f\|_2^{\frac{2}{q}}.$$

We also require the level-d inequality of [6], which is a consequence (not immediate) of their hypercontractive inequality, showing that global functions have low weight on the first d levels.

Theorem 2.7 *There exists $C > 0$ such that if $n \geq 2^{Cd^3} \log^{Cd}\left(\frac{1}{\|f\|_2}\right)$ and $f: S_n \to \mathbb{Z}$ is $(2d, \epsilon)$-global then*

$$\|f^{\leq d}\|_2 \leq 2^{Cd^4} \|f\|_2 \epsilon \log^{Cd}\left(\frac{1}{\|f\|_2^2}\right).$$

2.6 Eigenvalues of global Cayley graphs

We show the following stronger version of Theorem 1.9. Let $\mathrm{T}: L^2(S_n) \to L^2(S_n)$ be an operator that commutes with the action of S_n either from the left or from the right. Then we write

$$r_d(\mathrm{T}) = \sup_{f \in V_d} \frac{\|\mathrm{T}f\|_2}{\|f\|_2} \quad \text{and} \quad r_\lambda(\mathrm{T}) = \sup_{f \in V_\lambda} \frac{\|\mathrm{T}f\|_2}{\|f\|_2}.$$

When T is self-adjoint $r_d(\mathrm{T})$ is the largest eigenvalue of T inside V_d; in general, $r_d(\mathrm{T})^2$ is the largest eigenvalue of $\mathrm{T}^*\mathrm{T}$. Similarly for r_λ. Theorem 1.9 is immediate from the following result applied with $f = 1_A$ and $\epsilon = \sqrt{K\mu(A)}$, noting that $\|f\|_2 = \sqrt{\mu(A)}$ and $\mathrm{L}_f = \mu(A)T_A$, where T_A is the random walk operator corresponding to A.

Theorem 2.8 *There is an absolute constant C such that if $f: S_n \to \mathbb{Z}$ is $(2d, \epsilon)$-global for some $\epsilon > 0$ and $n \geq 2^{Cd^3} \log^{Cd}\left(\frac{1}{\|f\|_2}\right)$ then*

$$r_d(\mathrm{L}_f) \leq \frac{2^{Cd^4} \|f\|_2 \epsilon \log^{Cd}\left(\frac{1}{\|f\|_2^2}\right)}{n^{d/2}}.$$

If moreover f is a class function then

$$r_d(\mathrm{L}_f) \leq \frac{2^{Cd^4} \|f\|_2 \epsilon \log^{Cd}\left(\frac{1}{\|f\|_2^2}\right)}{n^{d}}.$$

To prove Theorem 2.8 we rely on the following lemma.

Lemma 2.9 *For each d and λ, the operator L_f agrees with $\mathrm{L}_{f^{=d}}$ on $V_{=d}$ and with $\mathrm{L}_{f^{=\lambda}}$ on $V_{=\lambda}$.*

Proof Let R_σ be the operator sending g to $g\sigma$. Then R_σ commutes with the action of S_n from the left. By Lemma 2.3 it therefore preserves the spaces $V_{=d}$ and $V_{=\lambda}$. Let $g \in V_{=d}$ and let $\sigma \in S_n$. Then as $R_\sigma g$ is in $V_{=d}$ and as $f^{=d}$ is the projection of f onto $V_{=d}$ we have

$$L_f(g)(\sigma) = \langle f, R_\sigma g\rangle = \left\langle f^{=d}, R_\sigma g\right\rangle = \mathrm{L}_{f^{=d}} g(\sigma).$$

The proof that L_f agrees with $\mathrm{L}_{f^{=\lambda}}$ on $V_{=\lambda}$ is similar. □

Proof of Theorem 2.8 Let $f : S_n \to \mathbb{Z}$ be $(2d, \epsilon)$-global.

The trace of the operator $\mathrm{L}^*_{f^{=d}} \mathrm{L}_{f^{=d}}$ is $\|f^{=d}\|_2^2$ by Lemma 2.4. By Lemma 2.9, and standard linear algebra we have

$$r_d\left(\mathrm{L}_f\right) = r_d\left(\mathrm{L}_{f^{=d}}\right) = \sqrt{r_d\left(\mathrm{L}^*_{f^{=d}} \mathrm{L}_{f^{=d}}\right)}.$$

On the other hand, the trace of a self-adjoint operator is the sum of its eigenvalues. Applying Lemma 2.3 gives

$$\min_{\lambda:\, d_\lambda = d} \dim(\lambda) \cdot r_d\left(\mathrm{L}^*_{f^{=d}} \mathrm{L}_{f^{=d}}\right) \le \mathrm{tr}\left(\mathrm{L}^*_{f^{=d}} \mathrm{L}_{f^{=d}}\right) = \|f^{=d}\|_2^2.$$

Putting everything together, plugging in Theorem 2.7 and Lemma 2.3 we obtain

$$r_d\left(L_f\right) \le \frac{\|f^{=d}\|_2}{\sqrt{\min_{\lambda:\, d_\lambda = d} \dim(\lambda)}} \le \frac{(ed)^{d/2} 2^{Cd^4} \|f\|_2 \epsilon \log^{Cd}\left(\frac{1}{\|f\|_2}\right)}{n^{d/2}},$$

for some absolute constant C. This implies that the theorem holds with $2C$ replacing C. When f is a class function the same proof works with $\dim(\lambda)$ replaced by $\dim(\lambda)^2$ to give

$$r_d\left(\mathrm{L}_f\right) \le \frac{2^{2Cd^4} \|f\|_2 \epsilon \log^{Cd}\left(\frac{1}{\|f\|_2}\right)}{n^d}. \qquad \square$$

2.7 The trace bound in high dimensions

The above upper bound on $r_d\left(\mathrm{L}_f\right)$ is complemented by the following simpler bound that is more effective when $\tilde{d}_\lambda$ is large.

Lemma 2.10 *Let $d < \frac{n}{10}$ and suppose that $\tilde{d}_\lambda \ge d$. Then $r_\lambda\left(\mathrm{L}_f\right) \le \left(\frac{ed}{n}\right)^{d/2} \|f\|_2$.*

Proof The trace of $\mathrm{L}^*_f \mathrm{L}_f$ is $\|f\|_2^2$. On the other hand, by Lemma 2.3

$$\|f\|_2^2 = \mathrm{tr}\left(\mathrm{L}^*_f \mathrm{L}_f\right) \ge \dim(\lambda)\, r_\lambda\left(\mathrm{L}_f\right)^2 \ge \left(\frac{n}{ed}\right)^d r_\lambda\left(\mathrm{L}_f\right)^2. \qquad \square$$

2.8 Proof of our 1% stability results

Here we prove a version of Theorem 1.5 that is stronger in two ways: we consider any triple of global sets with density $\ge n^{-O(1)}$ and we establish the product mixing phenomenon. For a function $f : S_n \to \mathbb{R}$ we write $\tilde{f}$ for $f \cdot \mathrm{sign}$. Theorem 1.5 (in contrapositive form) is immediate from the following by setting $f = g = h = 1_A$, noting that $\tilde{f} = f$ if f is supported on A_n.

Theorem 2.11 *Fix $C > 0$ and suppose that n is sufficiently large. Let $f, g, h\colon S_n \to \{0,1\}$ with $\mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h] \geq n^{-C}$ be $\left(i, n^{\frac{i}{4}}\right)$-relatively global for each $i \leq 8\lceil C\rceil$. Then*

$$\mathbb{E}_{\sigma,\tau\sim S_n}[f(\sigma)g(\tau)h(\sigma\tau)] = \mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h]\left(1 \pm n^{-0.1}\right) + \mathbb{E}\left[\tilde{f}\right]\mathbb{E}\left[\tilde{g}\right]\mathbb{E}\left[\tilde{h}\right].$$

The first step of the proof is to separate the left hand side $\mathbb{E}_{\sigma,\tau\sim S_n}[f(\sigma)g(\tau) \times h(\sigma\tau)]$ into low degree terms and high degree ones, as in the following lemma.

Lemma 2.12 *Let $d < \frac{n}{2} - 1$. Then*

$$\mathbb{E}_{\sigma,\tau\sim S_n}[f(\sigma)g(\tau)h(\sigma\tau)] = \sum_{i=0}^{d}\left\langle g^{=i}, \mathrm{L}_f h^{=i}\right\rangle + \sum_{i=0}^{d}\left\langle \tilde{g}^{=i}, \mathrm{L}_{\tilde{f}}\tilde{h}^{=i}\right\rangle + \sum_{\lambda\colon \tilde{d}_\lambda > d}\left\langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda}\right\rangle.$$

Proof As L_f commutes with the action of S_n from the right it preserves each $V_{=\lambda}$. We may therefore use orthogonality to obtain the following expansion into isotypical parts:

$$\begin{aligned}\mathbb{E}_{\sigma,\tau\sim S_n}[f(\sigma)g(\tau)h(\sigma\tau)] &= \left\langle g, \mathrm{L}_f h\right\rangle = \sum_{\lambda\vdash n}\left\langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda}\right\rangle \\ &= \sum_{\lambda\colon \tilde{d}_\lambda \leq d}\left\langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda}\right\rangle + \sum_{\lambda\colon \tilde{d}_\lambda > d}\left\langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda}\right\rangle.\end{aligned}$$

Regrouping terms of small degree. As $d < \frac{n}{2} - 1$ at most one of d_{λ^t}, d_λ can be at most d, so

$$\sum_{\tilde{d}_\lambda \leq d}\left\langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda}\right\rangle = \sum_{\lambda\colon d_\lambda \leq d}\left\langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda}\right\rangle + \sum_{\lambda\colon d_\lambda \leq d}\left\langle g^{=\lambda^t}, \mathrm{L}_f h^{=\lambda^t}\right\rangle.$$

As $V_{=i} = \sum_{\lambda\colon d_\lambda = i} V_{=\lambda}$ we have

$$\sum_{\lambda\colon d_\lambda \leq d}\left\langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda}\right\rangle = \sum_{i=0}^{d}\left\langle g^{=i}, \mathrm{L}_f h^{=i}\right\rangle.$$

Regrouping terms of small 'dual' degree. On the other hand, by Lemma 2.3 we have $g^{=\lambda^t} = \widetilde{\tilde{g}^{=\lambda}}$. Hence,

$$\begin{aligned}\sum_{d_\lambda \leq d}\left\langle g^{=\lambda^t}, \mathrm{L}_f h^{=\lambda^t}\right\rangle &= \sum_{d_\lambda \leq d}\left\langle \widetilde{\tilde{g}^{=\lambda}}, \mathrm{L}_f\left(\widetilde{\tilde{h}^{=\lambda}}\right)\right\rangle \\ &= \sum_{d_\lambda \leq d}\mathbb{E}_{\pi,\sigma}\left[\tilde{f}(\sigma)\tilde{g}^{=\lambda}(\pi)\tilde{h}^{=\lambda}(\sigma\pi)\right]\end{aligned}$$

$$= \sum_{d_\lambda \le d} \left\langle \mathrm{L}_{\tilde{f}} \tilde{h}^{=\lambda}, \tilde{g}^{=\lambda} \right\rangle$$

$$= \sum_{i=0}^{d} \left\langle \mathrm{L}_{\tilde{f}} \tilde{h}^{=i}, \tilde{g}^{=i} \right\rangle. \qquad \square$$

We now prove a lemma showing that the high degree terms are negligible.

Lemma 2.13

$$\left| \sum_{\lambda:\, \tilde{d_\lambda} > d} \langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda} \rangle \right| \le \left(\frac{n}{e(d+1)} \right)^{-\frac{d+1}{2}} \|f\|_2 \|g\|_2 \|h\|_2.$$

Proof By Lemma 2.10 and Cauchy–Schwarz we have

$$\sum_{\lambda:\, \tilde{d_\lambda} > d} \langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda} \rangle \le \sum_{\lambda:\, \tilde{d_\lambda} > d} r_\lambda \left(\mathrm{L}_f\right) \|h^{=\lambda}\|_2 \|g^{=\lambda}\|_2$$

$$\le \left(\frac{n}{e(d+1)} \right)^{-\frac{d+1}{2}} \|f\|_2 \sum_{\lambda \vdash n} \|h^{=\lambda}\|_2 \|g^{=\lambda}\|_2,$$

where $\sum_{\lambda \vdash n} \|h^{=\lambda}\|_2 \|g^{=\lambda}\|_2 \le \sqrt{\sum_{\lambda \vdash n} \|h^{=\lambda}\|_2^2 \sum_{\lambda \vdash n} \|g^{=\lambda}\|_2^2} = \|h\|_2 \|g\|_2.$ $\square$

We now prove the theorem by combining the bound on $\|f^{\le d}\|_2^2$ from the level-d inequality with the bounds from Lemma 2.13 on the eigenvalues of L_f that correspond to large degrees.

Proof of Theorem 2.11 Let $d = 4\lceil C \rceil$. We have

$$\langle \mathrm{L}_f g, h \rangle = \left\langle \mathrm{L}_f h^{=0}, g^{=0} \right\rangle + \left\langle \mathrm{L}_{\tilde{f}} \tilde{h}^{=0}, \tilde{g}^{=0} \right\rangle$$

$$\pm \sum_{i=1}^{d} \left| \left\langle g^{=i}, \mathrm{L}_f h^{=i} \right\rangle + \left\langle \tilde{g}^{=i}, \mathrm{L}_{\tilde{f}} \tilde{h}^{=i} \right\rangle \right|$$

$$\pm \sum_{\tilde{d_\lambda} > d} \left| \langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda} \rangle \right|.$$

The main terms are $\langle \mathrm{L}_f h^{=0}, g^{=0} \rangle = \mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h]$ and $\left\langle L_{\tilde{f}} \tilde{h}^{=0}, \tilde{g}^{=0} \right\rangle = \mathbb{E}\left[\tilde{f}\right]\mathbb{E}[\tilde{g}]\mathbb{E}\left[\tilde{h}\right]$.

By Lemma 2.13, using $\mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h] \ge n^{-C}$, we bound the high-degree error terms as

$$\sum_{\tilde{d_\lambda} > d} \left| \langle g^{=\lambda}, \mathrm{L}_f h^{=\lambda} \rangle \right| \le \left(\frac{n}{e(d+1)} \right)^{-\frac{d+1}{2}} \|f\|_2 \|g\|_2 \|h\|_2 \le \frac{\mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h]}{n}.$$

For any $i \le 8\lceil C \rceil$, as f is $(i, n^{i/4})$-relatively global it is $(i, n^{i/8}\|f\|_2)$-global, and similarly for g, h. By Theorems 2.7 (for g and h) and 2.8 (for f) we bound the low-degree error terms as

$$\sum_{i=1}^{d} \left|\left\langle \mathrm{L}_f h^{=i}, g^{=i} \right\rangle\right| \le \sum_{i=1}^{d} r_i \left(\mathrm{L}_f\right) \|g^{=i}\|_2 \|h^{=i}\|_2$$
$$\le \sum_{i=1}^{d} \frac{n^{3i/8}(\log n)^{O(1)}\mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h]}{n^{i/2}}$$
$$\le n^{-1/8}(\log n)^{O(1)}\mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h].$$

The same bounds hold replacing f, g, h on the left-hand side by $\tilde{f}, \tilde{g}, \tilde{h}$ (which have the same globalness properties) so the theorem follows. □

2.9 The linear terms dominate

For future reference, we conclude this section by noting that the above arguments establish the following key point that we mentioned in the introduction: when $\mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h]$ is large, the only significant contribution to $\mathbb{E}[f]\mathbb{E}[g]\mathbb{E}[h]$ comes from the linear terms. The following is immediate from Lemmas 2.12 and 2.13 with $d = 1$.

Proposition 2.14 *Let $f, g, h\colon A_n \to \{0, 1\}$ have densities α, β, γ in S_n. Then*

$$\left|\mathbb{E}_{\pi,\sigma\sim S_n}[f(\pi)g(\sigma\circ\pi)h(\sigma)] - 2\alpha\beta\gamma - 2\mathbb{E}_{\pi,\sigma\sim S_n}[f^{=1}(\pi)g^{=1}(\sigma\circ\pi)h^{=1}(\sigma)]\right|$$
$$\le \frac{2e}{n}\sqrt{\alpha\beta\gamma}.$$

3 Analysis of linear functions over the symmetric group

The main result of this section is our level-1 inequality for the pseudorandom part of a function. This in itself will suffice to recover Eberhard's result (for expository purposes we will give the argument at the end of the section). We will start by describing a canonical way to represent $f^{=1}$ as a linear combination of the dictators $x_{i\to j}$. This canonical representation $f^{=1} = \sum a_{ij}x_{i\to j}$ will naturally lead to a decomposition of $f^{=1}$ as a sum of its random part $f_{\mathrm{rand}} := \sum_{|a_{ij}|<\epsilon} a_{ij}\left(x_{i\to j} - \frac{1}{n}\right)$ and its structural part f_{struc}. Our level-1 inequality will bound $\|f_{\mathrm{rand}}\|_2$ by $\epsilon\|f\|_2$ up to logarithmic factors.

3.1 The normalized form of linear functions

We say that a linear function $\sum_{i,j} a_{ij}x_{i\to j}$ is in *normalized form* if for each i we have $\sum_{j=1}^{n} a_{ij} = 0$ and for each j we have $\sum_{i=1}^{n} a_{ij} = 0$. Every linear function in normalized form has zero expectation, i.e. is in $V_{=1}$. We will soon show the converse, i.e. that every $f \in V_{=1}$ has a normalized form, which is unique. First we give a simple formula for the inner product between two linear functions, which holds when at least one of them is in normalized form.

Lemma 3.1 *Let $f = \sum_{i,j\in[n]} a_{ij}x_{i\to j}$ be in normalized form. Let $g = \sum_{i,j} b_{ij}x_{i\to j}$ be an arbitrary linear function. Then*

$$\langle f, g\rangle = \frac{\sum_{i,j} a_{ij}b_{ij}}{n-1}.$$

Proof Consider the linear functionals $\varphi, \psi\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ given by

$$\varphi\left(\left(b_{ij}\right)_{i,j}\right) = \frac{1}{n-1}\sum_{i,j} a_{ij}b_{ij}$$

and

$$\psi\left(\left(b_{ij}\right)_{i,j}\right) = \left\langle f, \sum b_{ij}x_{i\to j}\right\rangle.$$

As both φ, ψ are linear it is enough to show that $\varphi = \psi$ on a basis. Hence, it is sufficient to prove the lemma when $g = x_{i\to j}$. There we may use the fact that f is in normalized form to deduce that

$$\begin{aligned}\langle f, g\rangle &= \frac{1}{n}a_{ij} + \frac{1}{n(n-1)}\sum_{i'\neq i, j'\neq j} a_{i'j'}\\ &= \frac{1}{n}a_{ij} - \frac{1}{n(n-1)}\sum_{j'\neq j} a_{ij'}\\ &= \frac{1}{n}a_{ij} + \frac{1}{n(n-1)}a_{ij}\\ &= \frac{1}{n-1}a_{ij}.\end{aligned}$$

This completes the proof. □

We are now ready to show that every function in $V_{=1}$ has a normalized form, which is unique. In fact, we give an explicit formula for the coefficients of each $f^{=1} \in V_{=1}$.

Lemma 3.2 *Let $f\colon S_n \to \mathbb{R}$. Let*

$$a_{ij} = \frac{n-1}{n}\left(\mathbb{E}\left[f_{i\to j}\right] - \mathbb{E}[f]\right).$$

Then $\sum_{i,j} a_{ij}x_{i\to j}$ is the unique normalized form of $f^{=1}$.

Proof First we note that $\sum_j \mathbb{E}[f_{i\to j}] = n\mathbb{E}f$ for each i, so $L := \sum_{i,j} a_{ij} x_{i\to j}$ is a linear function in normalized form. As $f^{=1}$ is the projection of f onto the space $V_{=1}$ of linear functions with expectation 0, to prove the lemma it suffices to show that L is the unique function in normalized form satisfying $\langle L, g\rangle = \langle f, g\rangle$ for all $g \in V_{=1}$.

By linearity, we can restrict attention to $g = x_{i\to j} - \frac{1}{n}$, as such functions g span $V_{=1}$. For such g we have

$$\langle f, g\rangle = \frac{1}{n}\mathbb{E}\left[f_{i\to j}\right] - \frac{1}{n}\mathbb{E}\left[f\right] = \frac{1}{n-1}a_{ij}.$$

On the other hand, as L is normalized form, by Lemma 3.1 we have

$$\langle L, g\rangle = \left\langle \sum_{i,j} a_{ij} x_{i\to j}, g\right\rangle = \frac{1}{n-1}a_{ij}.$$

Thus L is a normalised form of f. Furthermore, if $L' = \sum_{i,j} a'_{ij} x_{i\to j}$ is also a normalised form of f then the previous calculation gives $a_{ij} = (n-1)\langle L, g\rangle = (n-1)\left\langle L', g\right\rangle = a'_{ij}$, so we have uniqueness. □

For $f = \sum_{i,j} a_{ij} x_{i\to j}$ in normalized form, Lemma 3.1 gives the Parseval formula $\|f\|_2^2 = (n-1)^{-1}\sum_{i,j} a_{ij}^2$. For any linear function, not necessarily in normalized form, we still have the following upper bound, which has the same form up to a constant factor.

Lemma 3.3 *Let* $g = \sum_{i,j} a_{i,j}\left(x_{i\to j} - \frac{1}{n}\right)$ *be a function in* $V_{=1}$. *Then* $\|g\|_2^2 \le \frac{8}{n}\sum_{i,j} a_{ij}^2$.

Proof Computing, we get

$$\|g\|_2^2 = \sum_{i,j} a_{i,j}^2 \frac{1}{n}\left(1 - \frac{1}{n}\right) - \sum_{i,j}\sum_{i'\neq i} a_{i,j}a_{i',j}\frac{1}{n^2} - \sum_{i,j}\sum_{j'\neq j} a_{i,j}a_{i,j'}\frac{1}{n^2} + \sum_{i,j}\sum_{i'\neq i, j'\neq j} a_{i,j}a_{i',j'}\frac{1}{n^2(n-1)}.$$

We bound each term on the right hand side separately. Using $2|ab| \le a^2 + b^2$, each of the sums on the right hand side is $\le \frac{2}{n}\sum_{i,j} a_{i,j}^2$, and so $\|g\|_2^2 \le \frac{8}{n}\sum_{i,j} a_{i,j}^2$. □

Using Lemma 3.1 we can now derive a useful formula for the linear term in the count for the number of products in terms of the coefficient matrices of the normalised forms.

Lemma 3.4 *Let* $f = \sum_{i,j} a_{ij} x_{i\to j}$, $g = \sum_{i,j} b_{ij} x_{i\to j}$, $h = \sum_{i,j} c_{ij} x_{i\to j}$ *all be linear functions in normalized form. Let their coefficient matrices be defined by* $M_f = \left(a_{ij}\right)_{i,j}$, $M_g = \left(b_{ij}\right)_{i,j}$, *and* $M_h = \left(c_{ij}\right)_{i,j}$. *Then*

$$\mathbb{E}_{\sigma,\tau\sim S_n}\left[f(\sigma)g(\tau)h(\sigma\tau)\right] = \frac{1}{(n-1)^2}\left\langle M_g M_f, M_h\right\rangle.$$

Proof We have

$$R_\tau h(\sigma) = \sum_{i,j\in[n]} c_{ij} x_{i\to j}(\sigma\tau) = \sum_{i,j} c_{ij} \sum_k x_{i\to k}(\tau) x_{k\to j}(\sigma)$$
$$= \sum_{k,j} d_{kj}(\tau) x_{k\to j}(\sigma),$$

where $d_{kj}(\tau) = \sum_i c_{ij} x_{i\to k}(\tau)$. By Lemma 3.1 we therefore have

$$\mathrm{L}_f h(\tau) = \langle f, R_\tau h\rangle = \frac{1}{n-1} \sum_{i,j\in[n]} a_{ij} d_{ij}(\tau) = \frac{1}{n-1} \sum_{i,j,k} a_{ij} c_{kj} x_{k\to i}(\tau) =$$
$$= \frac{1}{n-1} \sum_{i,j} e_{ij} x_{i\to j}(\tau),$$

where $e_{ij} = \sum_k a_{jk} c_{ik}$. We deduce that

$$\mathbb{E}[f(\sigma) g(\tau) h(\sigma\tau)] = \langle g, \mathrm{L}_f h\rangle = \frac{1}{(n-1)^2} \sum_{i,j} e_{ij} b_{ij}$$
$$= \frac{1}{(n-1)^2} \sum_{i,j,k} b_{ij} a_{jk} c_{ik} = \frac{1}{(n-1)^2} \langle M_g M_f, M_h\rangle. \qquad \square$$

3.2 Global hypercontractivity and the level-1 inequality

The following lemma shows that Theorem 2.6 may be applied to linear functions with small coefficients. The lemma is applicable for f_{rand} defined above.

Lemma 3.5 *Let* $g = \sum_{i,j} a_{ij}\left(x_{i\to j} - \frac{1}{n}\right) \in V_{=1}$ *with* $|a_{ij}| < \epsilon$ *for all* i, j. *Then* g *is* $(2, \epsilon')$*-global for* $\epsilon' = 9\epsilon + \sqrt{\frac{8}{n-2}\sum_{i,j} a_{ij}^2}$.

Proof We need to show $\|g_{I\to J}\|_2 \le \epsilon'$ for any restriction with $|I| = |J| \le 2$. By averaging, it suffices to consider $|I| = |J| = 2$. Take distinct i, j, k, ℓ, and consider the restriction $i \to j, k \to \ell$, corresponding to the duumvirate $\mathcal{U}_{i\to j,k\to l}$. We apply the bound

$$\left\|g_{i\to j,k\to\ell}\right\|_2 \le |L| + \left\|\tilde{g}_{i\to j,k\to l}\right\|_2,$$

where we define $\tilde{g}\colon \mathcal{U}_{i\to j,k\to l} \to \mathbb{R}$ by

$$\tilde{g}(\pi) = \sum_{t\ne i,j} \sum_{q\ne k,\ell} a_{t,q}\left(x_{t\to q} - \frac{1}{n-2}\right),$$

and let

$$L = g_{i\to j,k\to l} - \tilde{g} = \left(1 - \frac{1}{n}\right)(a_{i,j} + a_{k,\ell}),$$

$$- \frac{1}{n} \sum_{t \neq j} a_{i,t} - \frac{1}{n} \sum_{t \neq \ell} a_{k,t} - \frac{1}{n} \sum_{t \neq i} a_{t,j} - \frac{1}{n} \sum_{t \neq k} a_{t,\ell}$$
$$+ \sum_{t \neq i,j} \sum_{q \neq k,\ell} a_{t,q} \left(\frac{1}{n-2} - \frac{1}{n} \right).$$

By the triangle inequality we have $|L| \le 9\epsilon$. For the second term above, we may use the aforementioned identification between $\mathcal{U}_{i\to j,k\to l}$ and S_{n-2} to apply Lemma 3.3, deducing that

$$\left\| \tilde{g}_{i\to j,k\to \ell} \right\|_2^2 \le \frac{8}{n-2} \sum_{t \neq i,j} \sum_{q \neq k,\ell} a_{t,q}^2.$$

Thus we obtain the required bound $\left\| g_{i\to j,k\to \ell} \right\|_2 \le \epsilon'$. □

3.3 A level-1 inequality for the pseudorandom part

We now show the desired upper bound on the L^2-norm $\| f_{\text{rand}} \|_2$ of the pseudorandom part of $f^{=1}$; in fact, we bound the right hand side of the bound $\| f_{\text{rand}} \|_2^2 \le 8X^2$ from Lemma 3.3, where X^2 is as in the following statement.

Lemma 3.6 *Let* $\epsilon \in \left(0, \frac{1}{2}\right)$ *and* $f\colon S_n \to \{0,1\}$ *with* $\mathbb{E}[f] \le \frac{1}{2}$. *Write* $f^{=1} = \sum_{i,j} a_{ij} x_{i\to j}$ *in normalized form and let* $f_{\text{rand}} = \sum_{i,j} a_{ij} \left(x_{i\to j} - \frac{1}{n} \right) 1_{|a_{ij}|<\epsilon}$. *Denote* $\epsilon'' = \max(\epsilon, \mathbb{E}[f])$. *Then*

$$X^2 := \frac{1}{n-1} \sum_{i,j} a_{ij}^2 1_{|a_{ij}|<\epsilon} \le \mathbb{E}[f] \epsilon'' \log^{O(1)} \left(\frac{1}{\epsilon''} \right).$$

Proof We note that $f_{\text{rand}} \in V_{=1}$ and by Lemma 3.1 we have

$$X^2 = \langle f_{\text{rand}}, f \rangle = \left\langle f_{\text{rand}}, f^{=1} \right\rangle.$$

Let $q = 10 \log\left(1/\epsilon''\right)$ and $\epsilon' = 9\epsilon + 3\| f_{\text{rand}} \|_2$, so that f_{rand} is $\left(2, \epsilon'\right)$-global by Lemma 3.5. Applying Hölder's inequality and Theorem 2.6, we obtain

$$\begin{aligned} X^2 = \langle f_{\text{rand}}, f \rangle &\le \| f_{\text{rand}} \|_q \, \| f \|_{q/(q-1)} \\ &\le q^{O(1)} (\epsilon')^{1-\frac{2}{q}} \, \| f_{\text{rand}} \|_2^{\frac{2}{q}} \, \| f \|_{q/(q-1)} \\ &= q^{O(1)} (\epsilon')^{1-\frac{2}{q}} \, \| f_{\text{rand}} \|_2^{\frac{2}{q}} \, \| f \|_2^{2-\frac{2}{q}}, \end{aligned}$$

where we used the fact that f is $\{0,1\}$-valued. Using $\| f_{\text{rand}} \|_2^2 \le 8X^2$ by Lemma 3.3 and rearranging we get

$$X^2 \le q^{O(1)} (\epsilon')^{\frac{q-2}{q-1}} \mathbb{E}[f].$$

We next consider two cases according to which of ϵ and X is larger.

1. If $X \le \epsilon$ then $\epsilon' \le 20\epsilon$, so we obtain

$$X^2 \le q^{O(1)} \epsilon^{\frac{q-2}{q-1}} \mathbb{E}[f].$$

2. If $X > \epsilon$ then $\epsilon' \le 20X$, so $X^2 \le q^{O(1)} X^{\frac{q-2}{q-1}} \mathbb{E}[f]$, yielding

$$X^2 \le q^{O(1)} \mathbb{E}[f]^{2-\frac{2}{q}}.$$

In both cases the lemma follows by plugging in $q = 10 \log\left(1/\epsilon''\right)$. □

3.4 Recovering Eberhard's result

For expository purposes, we will now cash in on our level-1 inequalities and deduce Eberhard's result (up to the polylog factor). We will repeatedly use the following upper bound on $|\langle MN, S\rangle|$ for three matrices M, N, S.

Lemma 3.7 *Let $M, N, S \in \mathbb{R}^{n\times n}$. Then we have*

$$|\langle MN, S\rangle| \le \|M\|_2 \|N\|_2 \|S\|_2.$$

Proof Write $Ne_i = v_i$, $Se_i = u_i$. By Cauchy–Schwarz

$$\begin{aligned} |\langle MN, S\rangle| &= \left| \sum_{i=1}^{n} \langle M v_i, u_i\rangle \right| \le \sum_{i=1}^{n} \|M\|_2 \|v_i\|_2 \|u_i\|_2 \\ &\le \|M\|_2 \sqrt{\sum_{i=1}^{n} \|v_i\|_2^2} \sqrt{\sum_{i=1}^{n} \|u_i\|_2^2} = \|M\|_2 \|N\|_2 \|S\|_2. \end{aligned}$$ □

Lemma 3.8 *Let $A, B, C \subseteq A_n$ and write $\alpha = \frac{|A|}{n!}$, $\beta = \frac{|B|}{n!}$, $\gamma = \frac{|C|}{n!}$. Then*

$$\begin{aligned} \Pr_{\sigma,\tau\sim S_n}[\sigma \in A, \tau \in B, \sigma\tau \in C] &\ge 2\alpha\beta\gamma - \frac{2e\sqrt{\alpha\beta\gamma}}{n} - \frac{\alpha\beta\gamma}{\sqrt{n}} \log^{O(1)}\left(\frac{1}{\alpha\beta\gamma}\right) \\ &- \alpha \log^{O(1)}\left(\frac{1}{\alpha}\right)\sqrt{\frac{\beta\gamma}{n}} - \beta \log^{O(1)}\left(\frac{1}{\beta}\right)\sqrt{\frac{\gamma\alpha}{n}} - \gamma \log^{O(1)}\left(\frac{1}{\gamma}\right)\sqrt{\frac{\alpha\beta}{n}}. \quad (2) \end{aligned}$$

Proof Write $f = 1_A$, $f^{=1} = \sum a_{ij} x_{i\to j}$, $\mathrm{A} = (a_{ij})$, $\mathrm{A}_- = \left(a_{ij} 1_{a_{ij}<0}\right)$, $\mathrm{A}_+ = \left(a_{ij} 1_{a_{ij}>0}\right)$. We use analogous notation for $g = 1_B$ and $h = 1_C$, with $g = \sum b_{ij} x_{i\to j}$ and $h = \sum c_{ij} x_{i\to j}$.

Then by Proposition 2.14

$$\begin{aligned} \Pr_{\sigma,\tau\sim A_n}[\sigma \in A, \tau \in B, \sigma\tau \in C] &= \mathbb{E}_{\sigma,\tau\sim S_n}[f(\sigma) g(\tau) h(\sigma\tau)] \\ \ge \quad & 2\alpha\beta\gamma + 2\mathbb{E}\left[f^{=1}(\sigma) g^{=1}(\tau) h^{=1}(\sigma\tau)\right] - \frac{2e\sqrt{\alpha\beta\gamma}}{n}. \end{aligned}$$

By Lemma 3.4 we have

$$\mathbb{E}\left[f^{=1}(\sigma)\,g^{=1}(\tau)\,h^{=1}(\sigma\tau)\right]=\frac{\langle \mathrm{BA},\mathrm{C}\rangle}{(n-1)^2}.$$

By our level 1 inequality (Lemma 3.6), applied with $\epsilon=\alpha$ (noting that each $a_{ij}>-\alpha$ by Lemma 3.2), we have

$$\|\mathrm{A}_-\|_2^2\leq n\alpha^2\log^{O(1)}(1/\alpha),$$

with analogous statements for B_- and C_-. By our Parseval lemma (Lemma 3.1) we have

$$\|\mathrm{A}_+\|_2^2\leq\|\mathrm{A}\|_2^2=(n-1)\,\alpha,$$

and similarly for B and C.

We may now expand $\langle \mathrm{BA},\mathrm{C}\rangle$ by writing $\mathrm{A}=\mathrm{A}_++\mathrm{A}_-$ and similarly for B and C. After discarding the terms with a positive contribution to $\langle \mathrm{BA},\mathrm{C}\rangle$ we are left with the four terms

$$\langle \mathrm{B}_+\mathrm{A}_-,\mathrm{C}_+\rangle,\langle \mathrm{B}_-\mathrm{A}_+,\mathrm{C}_-\rangle,\langle \mathrm{B}_+\mathrm{A}_+,\mathrm{C}_-\rangle,\langle \mathrm{B}_-\mathrm{A}_-,\mathrm{C}_-\rangle.$$

The lemma now follows from Lemma 3.7. □

When (A,B,C) is product-free the above immediately implies the following bounds on their densities, as if $\min(\alpha\beta,\beta\gamma,\gamma\alpha)\geq\frac{\log^R n}{n}$ for sufficiently large R then all terms bar $2\alpha\beta\gamma$ in the right hand side of (2) are $o(\alpha\beta\gamma)$.

Corollary 3.9 *Let (A,B,C) be product-free. Write $\alpha=\frac{|A|}{n!}$, $\beta=\frac{|B|}{n!}$ and $\gamma=\frac{|C|}{n!}$. Then*

$$\min(\alpha\beta,\beta\gamma,\gamma\alpha)\leq\frac{\log^{O(1)}n}{n}.$$

For future reference, we also note the following slightly stronger bound for the regime $\beta,\gamma\geq\epsilon>\frac{1}{\delta\sqrt{n}}$, where we can replace the factor $\log^{O(1)}n$ by $\frac{1}{\epsilon}\log^{O(1)}\left(\frac{1}{\epsilon}\right)$.

Corollary 3.10 *Let (A,B,C) be product-free. Write $\alpha=\frac{|A|}{n!}$, $\beta=\frac{|B|}{n!}$ and $\gamma=\frac{|C|}{n!}$. Suppose* $\min(\beta,\gamma)\geq\epsilon$ *with $1/2>\epsilon>\frac{1}{\delta\sqrt{n}}$. Then*

$$\alpha\leq\frac{1}{\epsilon n}\log^{O(1)}\left(\frac{1}{\epsilon}\right).$$

4 Dictatorial structure

In this section we will expose the dictatorial structure of product-free sets and triples that are not too sparse. Throughout the remainder of the paper we adopt the following notation. We let $A, B, C \subseteq S_n$ and $f = 1_A, g = 1_B, h = 1_C$. We write the linear parts in normalized form as

$$f^{=1} = \sum a_{ij} x_{i\to j}, \quad g^{=1} = \sum_{i,j} b_{ij} x_{i\to j}, \quad h^{=1} = \sum_{i,j} c_{ij} x_{i\to j}.$$

We write $\mathrm{A} = (a_{ij})$, $\mathrm{B} = (b_{ij})$ and $\mathrm{C} = (c_{ij})$, where the matrices are distinguished from the corresponding sets by the font (and also by context). Proposition 2.14 shows that in order to understand

$$\mathbb{E}_{\sigma,\tau\sim S_n}\left[f(\sigma)\, g(\tau)\, h(\sigma\tau)\right]$$

it is sufficient to understand the linear part

$$\mathbb{E}_{\sigma,\tau\sim S_n}\left[f^{=1}(\sigma)\, g^{=1}(\tau)\, h^{=1}(\sigma\tau)\right],$$

for which Lemma 3.4 gives the formula

$$\mathbb{E}_{\sigma,\tau\sim S_n}\left[f^{=1}(\sigma)\, g^{=1}(\tau)\, h^{=1}(\sigma\tau)\right] = \frac{1}{(n-1)^2}\langle \mathrm{BA}, \mathrm{C}\rangle .$$

We will decompose the matrix A (and similarly B, C) into three parts:

1. The matrix A_- contains the negative coefficients of A, and so represents the negative correlations that A has with dictators.
2. The matrix $\mathrm{A}_{\mathrm{rand}}$ contains the small positive coefficients of A, and so represents the pseudorandom part of $f^{=1}$.
3. The matrix $\mathrm{A}_{\mathrm{struc}}$ contains the large coefficients of A, and so corresponds to the dictators with which A is heavily correlated.

We expand $\langle \mathrm{BA}, \mathrm{C}\rangle$ according to this decomposition and show that most of the negative contributions come from the terms

$$\langle \mathrm{B}_{\mathrm{struc}}\mathrm{A}_-, \mathrm{C}_{\mathrm{struc}}\rangle + \langle \mathrm{B}_-\mathrm{A}_{\mathrm{struc}}, \mathrm{C}_{\mathrm{struc}}\rangle + \langle \mathrm{B}_{\mathrm{struc}}\mathrm{A}_{\mathrm{struc}}, \mathrm{C}_-\rangle ,$$

namely those compatible triples of dictators for which two of the matrices have a strong positive correlation and the third has a negative correlation.

4.1 Parameters

The following parameters will be used throughout the remainder of the paper. We let

$$\mu(A) = \alpha, \quad \mu(B) = \beta, \quad \mu(C) = \gamma, \quad \text{where} \quad \alpha\beta\gamma > n^{-O(1)}.$$

We fix R much larger than all the absolute constants implicitly appearing in our $O(1)$ notation and suppose that n is sufficiently large with respect to R. Motivated by the calculation in Lemma 4.2 below, we let

$$\delta = \log^{-R} n, \quad \epsilon_A = n\delta\alpha \min(\beta,\gamma), \quad \epsilon_B = n\delta\beta \min(\alpha,\gamma),$$

$$\epsilon_C = n\delta\gamma \min(\alpha, \beta).$$

By Corollary 3.9 and the definition of δ, we have $\epsilon_A, \epsilon_B, \epsilon_C = o(1)$ when (A, B, C) are product-free.

Note also that in our exact result, which is the case of most interest, we have $A = B = C$ and $\alpha = \beta = \gamma = \Theta\left(\frac{1}{\sqrt{n}}\right)$, so $\epsilon_A = \epsilon_B = \epsilon_C = \Theta(\delta)$.

4.2 Our decomposition

For a matrix $M = (m_{ij})$ and an interval $I \subseteq \mathbb{R}$ we write $M_I = (a_{ij} 1_{a_{ij} \in I})$. As mentioned above, our idea is to decompose our matrix A as the sum

$$\mathrm{A} = \mathrm{A}_- + \mathrm{A}_{\mathrm{struc}} + \mathrm{A}_{\mathrm{rand}},$$

where $\mathrm{A}_- = \mathrm{A}_{(-\infty,0)}$, $\mathrm{A}_{\mathrm{rand}} = \mathrm{A}_{(0,\epsilon_A)}$ and $\mathrm{A}_{\mathrm{struc}} = \mathrm{A}_{[\epsilon_A,\infty)}$. We decompose B and C similarly with ϵ_B and ϵ_C replacing ϵ_A.

As mentioned in the introduction, the key to our approach is to combine Lemma 3.7 with the following upper bounds on $\|\mathrm{A}_{\mathrm{rand}}\|_2^2$ and $\|\mathrm{A}_-\|_2^2$, which will follow easily from our level-1 inequalities in the previous section.

Lemma 4.1 *Let $A \subseteq S_n$ and let $\mathrm{A}_-, \mathrm{A}_{struc}, \mathrm{A}_{rand}$ as above. Then*

1. $\|\mathrm{A}_-\|_2^2 \leq n\alpha^2 \log^{O(1)}(1/\alpha)$.
2. $\|\mathrm{A}_{rand}\|_2^2 \leq n\alpha\epsilon_A \log^{O(1)}(1/\alpha)$.
3. $\|\mathrm{A}_{struc}\|_2^2 \leq n\alpha$.

Analogous statements hold for B and C.

Proof Statements (1) and (2) follow immediately from Lemma 3.6. Statement (3) follows from Lemma 3.1 and the fact that $\|f^{=1}\|_2^2 \leq \|f\|_2^2 = \alpha$. □

We are now ready to show that the only significant negative contributions to $\langle \mathrm{BA}, \mathrm{C}\rangle$ come from two structure matrices and one negative coefficient matrix.

Lemma 4.2 *If $\alpha\beta\gamma > n^{-O(1)}$ then*

$$\langle \mathrm{BA}, \mathrm{C}\rangle \geq \langle \mathrm{B}_{struc}\mathrm{A}_-, \mathrm{C}_{struc}\rangle + \langle \mathrm{B}_-\mathrm{A}_{struc}, \mathrm{C}_{struc}\rangle + \langle \mathrm{B}_{struc}\mathrm{A}_{struc}, \mathrm{C}_-\rangle + \sqrt{\delta} n^2 (\log n)^{O(1)} \alpha\beta\gamma.$$

Proof We expand the left hand side according to the decomposition $\mathrm{A} = \mathrm{A}_- + \mathrm{A}_{\mathrm{struc}} + \mathrm{A}_{\mathrm{rand}}$ and similarly for B, C. For a lower bound we can discard all the terms that involve an even number of $\mathrm{A}_-, \mathrm{B}_-, \mathrm{C}_-$ as those have a non-negative contribution to $\langle \mathrm{BA}, \mathrm{C}\rangle$. For the remaining terms not listed on the right hand side above we may apply Lemmas 3.7 and 4.1 to deduce that they have absolute value at most

$$< n^{1.5} (\log n)^{O(1)} \left(\left(\sqrt{\epsilon_B} + \sqrt{\epsilon_C}\right) \alpha \sqrt{\beta\gamma}\right)$$

$$
\begin{aligned}
&+ n^{1.5}(\log n)^{O(1)} \left(\left(\sqrt{\epsilon_A} + \sqrt{\epsilon_C}\right) \beta \sqrt{\alpha\gamma}\right) \\
&+ n^{1.5}(\log n)^{O(1)} \left(\left(\sqrt{\epsilon_A} + \sqrt{\epsilon_B}\right) \gamma \sqrt{\alpha\beta}\right) \\
&+ \alpha\beta\gamma n^{1.5}(\log n)^{O(1)} \\
\leq\ & \sqrt{\delta} n^2 (\log n)^{O(1)} \alpha\beta\gamma.
\end{aligned}
$$

Here, the first three lines corresponds to terms such as $\langle \mathrm{B_{rand}A_-}, \mathrm{C_{struc}} \rangle$ and $\langle \mathrm{B_{rand}A_-}, \mathrm{C_{rand}} \rangle$, while the fourth line corresponds to the term $\langle \mathrm{B_-A_-}, \mathrm{C_-} \rangle$. The final inequality follows by plugging in the values of $\epsilon_A, \epsilon_B, \epsilon_C$. □

5 Star structure

In this section we will refine the dictatorial structure established in the previous section to extract a strong star structure that explains how some set or triple in A_n can be quite dense yet product-free. These stability results will then be refined by bootstrapping arguments in the next section to deduce our exact and strong stability results.

Equivalence and inversion. The equation $ab = c$ can be written in 6 equivalent ways, e.g. we may write $ca^{-1} = b$ and $b^{-1}a^{-1} = c^{-1}$. Thus if the triple (A, B, C) is product-free then we have 6 equivalent product-free triples such as $\left(C, A^{-1}, B\right)$ and $\left(B^{-1}, A^{-1}, C^{-1}\right)$. The structure explaining this product-freeness may appear in any of 6 different forms, so to avoid cumbersome statements, we will say that a certain structural statement for (A, B, C) holds *up to equivalence* if it holds when (A, B, C) is replaced by one of its 6 equivalent triples. Similarly, for a single product-free set A, the structural statement for A may apply to A or A^{-1}, so we will say that it holds *up to inversion*.

5.1 Goals of this section

Our first main result of this section will show that any product-free set has a strong star structure under a fairly mild assumption on its density (recall that $\delta^{-1} = \log^R n$).

Proposition 5.1 *Suppose that A is product-free with $\mu(A) \geq \delta^{-2} n^{-2/3}$. Then up to inversion there exist $x \in [n]$ and $I \subseteq [n]$ such that*

$$\mu\left(A \setminus 1_{x \to I}\right) \leq O(\delta^{-2}) n^{-\frac{2}{3}}.$$

Moreover, for each $i \in I$ we have $\mu\left(A_{x \to i}\right) \geq n^{-1/3}$.

Our second main result of the section describes the star structure for product-free triples under mild density assumptions: up to equivalence B and C must be strongly correlated with stars at some common vertex x.

Proposition 5.2 *Suppose that (A, B, C) is product-free with*

$$\alpha \min(\beta, \gamma)^2, \beta \min(\alpha, \gamma)^2, \gamma \min(\alpha, \beta)^2 \geq \delta^{-6} n^{-2}.$$

Then up to equivalence there exist $x \in [n]$ *and* $I, J \subseteq [n]$ *such that the following hold*:

1. *We have* $\mu_B(1_{x\to I}) \geq \frac{1}{100}$ *and* $\mu_C(1_{x\to J}) \geq \frac{1}{100}$.
2. *For each* $i \in I$ *and* $j \in J$ *we have* $\mu(B_{x\to i}) \geq \epsilon_B$ *and* $\mu\left(C_{x\to j}\right) \geq \epsilon_C$.

5.2 Associated stars

As mentioned at the start of Sect. 4, the notation introduced there (such as a_{ij} and ϵ_A) will be in force for the remainder of the paper. We now introduce some further notation, that will also be used throughout the remainder of the paper, to describe the stars associated to the structured parts of A, B and C.

For each $i \in [n]$, we define the *associated star* for (A, i) by

$$S_A(i) = \bigcup_{j\in L_A(i)} 1_{i\to j}, \qquad \text{where } L_A(i) = \{j : a_{ij} > \epsilon_A\}.$$

Similarly, we define the *associated inverse star* for (A, i) by

$$S'_A(i) = \bigcup_{j\in L'_A(i)} 1_{j\to i}, \qquad \text{where } L'_A(i) = \{j : a_{ji} > \epsilon_A\}.$$

We write

$$s_A(i) = \frac{\sum_{j\in L_i(A)} a_{ij}}{n-1} \quad \text{and} \quad s'_A(i) = \frac{\sum_{j\in L'_A(i)} a_{ji}}{n-1}.$$

We may interpret $s_A(i)$ and $s'_A(i)$ combinatorially as the correlation between A and the corresponding associated (inverse) star by noting that

$$s_A(i) = \mu(A \cap S_A(i)) - \mu(A)\,\mu(S_A(i))$$

and similarly for $s'_A(i)$. We also define the corresponding notions for B and C similarly. We say that an associated star $S_A(i)$ is *small* if $s_A(i) \leq \delta\mu(A)$, or otherwise we say that it is *large*. We use this terminology for associated stars and associated inverse stars for each of A, B, C.

5.3 Overview of proof

We will now give an overview of the arguments used to extract star structure from the inequalities discussed above. For simplicity in the overview we will concentrate on the case of a single product-free set A, which is analysed by applying the inequalities with $A = B = C$.

For terms such as $\langle \mathrm{A}_{\mathrm{struc}}\mathrm{A}_-, \mathrm{A}_{\mathrm{struc}}\rangle$ we use the fact that each coefficient of A_- is at least $-\alpha$, which provides the following lower bound in terms of the associated stars:

$$\frac{1}{(n-1)^2}\langle \mathrm{A}_{\mathrm{struc}}\mathrm{A}_-, \mathrm{A}_{\mathrm{struc}}\rangle \geq -\alpha \sum_{i=1}^{n} s_A(i)^2.$$

Summing over all the similar terms, we thus reduce our goal to a lower bound

$$-\alpha \sum_{i=1}^{n} \left(s_A(i)^2 + s_A(i)\, s'_A(i) + s'_A(i)^2 \right) \geq -\alpha^3 + X, \tag{3}$$

where X dominates the error terms unless A has strong star structure.

There will be two main ingredients in this bound. The first ingredient concerns getting rid of the small associated stars: we use the level-1 inequality (Lemma 3.6) to show that the ith contribution to (3) is negligible unless either $S_A(i)$ or $S'_A(i)$ is large.

The second ingredient provides a combinatorial analysis of large associated (inverse) stars, which is motivated by the heuristic that such stars should be essentially disjoint. Writing $s_A := \sum_i s_A(i)$ and $s'_A = \sum_i s'_A(i)$, we should therefore expect $s_A + s'_A$ to be essentially bounded by α. In terms of the rescaled star sizes

$$v_i := \frac{s_A(i)}{s_A + s'_A} \quad \text{and} \quad v'_i := \frac{s'_A(i)}{s_A + s'_A},$$

our goal is then essentially reduced to showing that

$$\sum_{i=1}^{n} v_i^2 + v_i v'_i + v_i'^2 \leq 1,$$

and moreover that if the sum is close to 1 then some v_i or v'_i is close to 1, which is equivalent to the required approximation of A by an associated star or inverse star.

5.4 Relating significant negative contributions to associated stars

We start the implementation of the above overview by expressing the significant negative contributions to $\langle \mathrm{BA}, \mathrm{C} \rangle$ (as described in Lemma 4.2) in terms of associated stars.

We write $\overrightarrow{1}$ for the all-ones vector in $\mathbb{R}^n$. The following inequalities are immediate from the facts that each coefficient A_- is at least $-\alpha$, and similarly for B and C, using the identity $\left\langle X\overrightarrow{1}, Y\overrightarrow{1} \right\rangle = \sum_{i,j,k} x_{ij} y_{ik}$ for any matrices $X = (x_{ij})$, $Y = (y_{ij})$.

Lemma 5.3 *We have the following inequalities.*

$$\frac{1}{(n-1)^2} \langle \mathrm{B}_- \mathrm{A}_{struc}, \mathrm{C}_{struc} \rangle \geq -\frac{\beta}{(n-1)^2} \left\langle \mathrm{A}^t_{struc} \overrightarrow{1}, \mathrm{C}^t_{struc} \overrightarrow{1} \right\rangle = -\beta \sum_{i=1}^{n} s'_A(i)\, s'_C(i)$$

$$\frac{1}{(n-1)^2} \langle \mathrm{B}_{struc} \mathrm{A}_{struc}, \mathrm{C}_- \rangle \geq \frac{-\gamma}{(n-1)^2} \left\langle \mathrm{A}_{struc} \overrightarrow{1}, \mathrm{B}^t_{struc} \overrightarrow{1} \right\rangle = -\gamma \sum_{i=1} s_A(i)\, s'_B(i)$$

$$\frac{1}{(n-1)^2} \langle \mathrm{B}_{struc} \mathrm{A}_-, \mathrm{C}_{struc} \rangle \geq \frac{-\alpha}{(n-1)^2} \left\langle \mathrm{B}_{struc} \overrightarrow{1}, \mathrm{C}_{struc} \overrightarrow{1} \right\rangle = -\alpha \sum_{i=1} s_B(i)\, s_C(i)$$

5.5 Small associated stars have a small contribution

We will now bound the terms appearing in Lemma 5.3 by replacing the matrices such as $\mathrm{A}_{\mathrm{struc}}$ that represent the dictatorial structure of A with certain matrices $\mathrm{A}_{\star}$ that represent the star structure of A.

We let $\mathrm{A}_{\star}$ be obtained from $\mathrm{A}_{\mathrm{struc}}$ as follows. Each row $e_i\mathrm{A}_{\mathrm{struc}}$ of $\mathrm{A}_{\mathrm{struc}}$ either corresponds to a large associated star if the sum of its coefficients is $> \delta\mu(A)(n-1)$, or otherwise it corresponds to a small associated star. We let $\mathrm{A}_{\star}$ be obtained from $\mathrm{A}_{\mathrm{struc}}$ by replacing all the rows that correspond to small associated stars with 0.

Similarly, we let $\mathrm{A}'_{\star}$ corresponding to large associated inverse stars be obtained from $\mathrm{A}^t_{\mathrm{struc}}$ by replacing its rows that correspond to small associated inverse stars with 0. We define $\mathrm{B}_{\star}$, $\mathrm{C}_{\star}$, $\mathrm{B}'_{\star}$ and $\mathrm{C}'_{\star}$ similarly.

The following lemma will replace the terms $\left\langle \mathrm{A}^t_{\mathrm{struc}}\overrightarrow{1}, \mathrm{C}^t_{\mathrm{struc}}\overrightarrow{1}\right\rangle$, $\left\langle \mathrm{A}_{\mathrm{struc}}\overrightarrow{1}, \mathrm{B}^t_{\mathrm{struc}}\overrightarrow{1}\right\rangle$, and $\left\langle \mathrm{B}_{\mathrm{struc}}\overrightarrow{1}, \mathrm{C}_{\mathrm{struc}}\overrightarrow{1}\right\rangle$ by the corresponding terms $\left\langle \mathrm{A}'_{\star}\overrightarrow{1}, \mathrm{C}'_{\star}\overrightarrow{1}\right\rangle$, $\left\langle \mathrm{A}_{\star}\overrightarrow{1}, \mathrm{B}'_{\star}\overrightarrow{1}\right\rangle$, and $\left\langle \mathrm{B}_{\star}\overrightarrow{1}, \mathrm{C}_{\star}\overrightarrow{1}\right\rangle$. The applicability of the lemma to these terms will follow from the level-1 inequality (Lemma 3.6).

Lemma 5.4 *Let* $M = \left(m_{ij}\right), N = \left(n_{ij}\right)$ *be matrices with entries in* $[n^{-2}, 1]$. *Let* $\eta_1, \eta_2 > 0$, *and suppose for each* i *that*

$$\|\mathrm{M}_{(2^{-i},2^{1-i}]}\|_2^2 \le \eta_1 2^{-i}\log^{O(1)} n,$$

$$\|\mathrm{N}_{(2^{-i},2^{1-i}]}\|_2^2 \le \eta_2 2^{-i}\log^{O(1)} n.$$

Define matrices $M_{\star} = (m^{\star}_{ij})$ *and* $N_{\star} = (n^{\star}_{ij})$ *by*

$$m^{\star}_{ij} = m_{ij}1_{\sum_{j'} m_{ij'} > \delta\eta_1} \quad \text{and} \quad n^{\star}_{ij} = n_{ij}1_{\sum_{j'} n_{ij'} > \delta\eta_2}.$$

Then we have

$$\left|\left\langle M\overrightarrow{1}, N\overrightarrow{1}\right\rangle - \left\langle M_{\star}\overrightarrow{1}, N_{\star}\overrightarrow{1}\right\rangle\right| \le \delta\eta_1\eta_2\log^{O(1)} n + \delta\eta_1\|N_{\star}\|_1 + \delta\eta_2\|M_{\star}\|_1.$$

Proof We have

$$\left|\left\langle (M - M_{\star})\overrightarrow{1}, N_{\star}\overrightarrow{1}\right\rangle\right| \le \|(M - M_{\star})\overrightarrow{1}\|_{\infty}\|N_{\star}\overrightarrow{1}\|_1 \le \delta\eta_1\|N_{\star}\|_1,$$

and similarly

$$\left|\left\langle M_{\star}\overrightarrow{1}, (N - N_{\star})\overrightarrow{1}\right\rangle\right| \le \delta\eta_2\|M_{\star}\|_1.$$

Let $\tilde{M} = M - M_{\star}$ and $\tilde{N} = N - N_{\star}$. It remains to show that

$$\left\langle \tilde{M}\overrightarrow{1}, \tilde{N}\overrightarrow{1}\right\rangle \le \delta\eta_1\eta_2\log^{O(1)} n.$$

We will use the dyadic expansion

$$\tilde{M}=\sum_{i=1}^{2\lceil\log n\rceil}\tilde{M}_{(2^{-i},2^{1-i}]}\quad\text{and}\quad\tilde{N}=\sum_{i=1}^{2\lceil\log n\rceil}\tilde{N}_{(2^{-i},2^{1-i}]}.$$

Each of the resulting terms will be bounded using the following claim.

Claim 5.5 *Let M', N' be matrices such that each row of M' has at most m' nonzero coefficients and each row of N' has at most n' nonzero coefficients. Then*

$$\left|\left\langle M'\overrightarrow{1}, N'\overrightarrow{1}\right\rangle\right| \leq \sqrt{m'n'}\|M'\|_2\|N'\|_2.$$

Proof This follows from Cauchy–Schwarz as

$$\begin{aligned}\left|\left\langle M'\overrightarrow{1}, N'\overrightarrow{1}\right\rangle\right| &= \left|\sum_i\left\langle \left(M'\right)^t e_i, \overrightarrow{1}\right\rangle\left\langle \left(N'\right)^t e_i, \overrightarrow{1}\right\rangle\right|\\ &\leq \sum_i \sqrt{m'}\left\|\left(M'\right)^t e_i\right\|_2 \sqrt{n'}\left\|\left(N'\right)^t e_i\right\|_2\\ &\leq \sqrt{m'n'}\left\|M'\right\|_2\left\|N'\right\|_2.\end{aligned}$$

□

For each row v of $\tilde{M}_{(2^{-i},2^{1-i}]}$, all nonzero coefficients are $\geq 2^{-i}$ and $\langle v, 1\rangle \leq \delta\eta_1$, so v has $\leq 2^i\delta\eta_1$ nonzero coefficients. Similarly each row of $\tilde{N}_{(2^{-j},2^{1-j}]}$ has $\leq 2^j\delta\eta_2$ nonzero coefficients. Applying the claim, we obtain

$$\begin{aligned}\left\langle \tilde{M}\overrightarrow{1}, \tilde{N}\overrightarrow{1}\right\rangle &= \sum_{i=1}^{2\lceil\log n\rceil}\sum_{j=1}^{2\lceil\log n\rceil}\left\langle \tilde{M}_{(2^{-i},2^{1-i}]}\overrightarrow{1}, \tilde{N}_{(2^{-j},2^{1-j}]}\overrightarrow{1}\right\rangle\\ &\leq \sum_{i=1}^{2\lceil\log n\rceil}\sum_{j=1}^{2\lceil\log n\rceil} 2^{\frac{i+j}{2}}\delta\sqrt{\eta_1\eta_2}\|\tilde{M}_{(2^{-i},2^{1-i}]}\|_2\|\tilde{N}_{(2^{-j},2^{1-j}]}\|_2\\ &\leq 4\delta\eta_1\eta_2\log^{O(1)} n,\end{aligned}$$

where the final inequality uses the assumption of the lemma. □

5.6 Combinatorial analysis of the star structure matrices

In the next section we will use Lemma 5.4 to complete our transition from dictatorial structure matrices $\mathrm{A}_{\mathrm{struc}}$ to the star-structure matrices $\mathrm{A}_\star$. To achieve this, we first need to control the error terms that will arise from applying Lemma 5.4, namely the L^1-norms of the star structure matrices.

As discussed in the overview, this comes down to a combinatorial argument showing that the corresponding large associated stars and inverse stars are essentially disjoint, provided that A, B, C (and so the parameters $\epsilon_A, \epsilon_B, \epsilon_C$) are sufficiently large.

This is captured by saying for each $E \in \{A, B, C\}$ that $\sum \mu(E \cap S) \approx \mu(E)$, where the sum is over all large associated stars and inverse stars of E.

Lemma 5.6 *Let $\epsilon > 0, \delta > 0$ and $E \subseteq S_n$ with $\mu(E) < 10^{-3}\epsilon^2$. Let $\mathcal{S}$ be a collection of stars and inverse stars, containing for each $x \in [n]$ at most one star at x and at most one inverse star at x. Suppose for each $S \in \mathcal{S}$ that $\delta\mu(E) \leq \mu(E \cap S)$ and $\frac{\epsilon}{2}\mu(S) \leq \mu(E \cap S)$. Then*

$$\sum_{S \in \mathcal{S}} \mu(E \cap S) \leq \mu(E) + \frac{25}{\epsilon^2}\mu(E)^2 + \frac{1.5}{\delta^2 n}.$$

Proof Write $\mathcal{S} = \mathcal{S}^+ \cup \mathcal{S}^-$, where $\mathcal{S}^+$ consists of stars and $\mathcal{S}^-$ consists of inverse stars.

We claim for any $\mathcal{S}' \subseteq \mathcal{S}^+$ or $\mathcal{S}' \subseteq \mathcal{S}^-$ that

$$\Sigma_{\mathcal{S}'} := \sum_{S \in \mathcal{S}'} \mu(E \cap S) \leq \mu(E) + \frac{5}{\epsilon^2}\Sigma_{\mathcal{S}'}^2. \tag{4}$$

To see this, we first apply Inclusion–Exclusion to get

$$\mu(E) \geq \Sigma_{\mathcal{S}'} - \sum_{\{S_1, S_2\} \subseteq \mathcal{S}'} \mu(S_1 \cap S_2). \tag{5}$$

Next, for any distinct stars $S_1, S_2 \in \mathcal{S}'$, we can write their intersection $S_1 \cap S_2$ as a union of duumvirates $U_{x_1 x_2 \to y_1 y_2}$, where $U_{x_1 \to y_1} \subseteq S_1$ and $U_{x_2 \to y_2} \subseteq S_2$, so

$$\mu(S_1 \cap S_2) \leq \left(\frac{n}{n-1}\right)\mu(S_1)\mu(S_2) \leq \frac{5}{\epsilon^2}\mu(E \cap S_1)\mu(E \cap S_2).$$

Summing over distinct S_1, S_2 in $\mathcal{S}'$ and substituting in (5) we deduce (4), as claimed.

Next we claim that $\Sigma_{\mathcal{S}^\pm} \leq 1.1\mu(E)$. Indeed, if this were false then we find $\mathcal{S}' \subseteq \mathcal{S}^\pm$ with $1.1\mu(E) \leq \Sigma_{\mathcal{S}'} \leq 2.1\mu(E)$. However, then (4) gives $1.1\mu(E) \leq \mu(E) + \frac{5}{\epsilon^2}(2.1\mu(E))^2$, which contradicts $\mu(E) < 10^{-3}\epsilon^2$. Thus the claim holds. As $\Sigma_{\mathcal{S}^\pm} \geq |\mathcal{S}^\pm|\delta\mu(E)$ we deduce $|\mathcal{S}^\pm| < 1.1\delta^{-1}$.

Now we repeat the calculation for (4) with $\mathcal{S}$ in place of $\mathcal{S}'$. The upper bound on $\mu(S_1 \cap S_2)$ needs an extra term of $\frac{1}{n}$ for a dictator $U_{x \to y}$ in the intersection of a star at x and an inverse star at y. As $|\mathcal{S}^\pm| < 1.1\delta^{-1}$, we obtain

$$\Sigma_{\mathcal{S}} := \sum_{S \in \mathcal{S}} \mu(E \cap S) \leq \mu(E) + \frac{5}{\epsilon^2}\Sigma_{\mathcal{S}}^2 + \frac{(1.1\delta^{-1})^2}{n}.$$

As $\Sigma_{\mathcal{S}} = \Sigma_{\mathcal{S}^+} + \Sigma_{\mathcal{S}^-} \leq 2.2\mu(E)$ the lemma follows. □

Lemma 5.6 can be restated in terms of the star-structure matrices. It translates to the following upper bound on their 1-norms.

Lemma 5.7 *For* (A, B, C) *as in Proposition 5.2 we have* $\frac{1}{n-1}\left(\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1\right) \le \alpha\,(1+2\delta)$, *etc for* B, C.

Proof Let $\mathcal{S}_A$ be the set of large associated stars for A. For each $S \in \mathcal{S}_A$ we have $\mu\,(A \cap S) \ge \delta\mu\,(A)$ by definition. On the other hand, for each dictator $1_{i\to j}$ contained in S we have

$$\epsilon_A < a_{ij} = \frac{n-1}{n}\left(\mu\left(A_{i\to j}\right) - \alpha\right),$$

which gives $\mu\left(A_{i\to j}\right) \ge \epsilon_A$, and so $\mu\,(A\cap S) \ge \epsilon_A\mu\,(S)$. Hence we can apply Lemma 5.6 with $\epsilon = \epsilon_A$, which gives

$$\frac{1}{n-1}\left(\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1\right) \le \sum_{S\in\mathcal{S}_A} \mu\,(A\cap S) \le \alpha + \frac{25\alpha^2}{\epsilon_A^2} + \frac{1.5}{\delta^2 n}.$$

Substituting $\epsilon_A = n\delta\alpha\min(\beta,\gamma)$ and using $\alpha\min\,(\beta,\gamma)^2 \ge \delta^{-6}n^{-2}$ gives $\epsilon_A^2/\alpha \ge \delta^{-4}$. As our assumptions imply $\alpha \ge \delta^{-3}n^{-1}$, we have $\frac{1.5}{\delta^2 n} \le 1.5\delta\alpha$, so the lemma follows. □

5.7 Reducing to large associated stars

The following lemma combines everything that we proved so far. It reduces us to upper bounding the star structure inner products $\left\langle \mathrm{A}'_\star\vec{1}, \mathrm{C}'_\star\vec{1}\right\rangle$, $\left\langle \mathrm{A}_\star\vec{1}, \mathrm{B}'_\star\vec{1}\right\rangle$ and $\left\langle \mathrm{B}_\star\vec{1}, \mathrm{C}_\star\vec{1}\right\rangle$.

Lemma 5.8 *Suppose that* (A, B, C) *are as in Proposition 5.2. Then we have*

$$\Pr_{\sigma,\tau\sim S_n}[\sigma\in A, \tau\in B, \sigma\tau\in C] \ge 2\alpha\beta\gamma(1-\delta^{1/4}) - \frac{2}{(n-1)^2}\left|\beta\left\langle \mathrm{A}'_\star\vec{1}, \mathrm{C}'_\star\vec{1}\right\rangle + \gamma\left\langle \mathrm{A}_\star\vec{1}, \mathrm{B}'_\star\vec{1}\right\rangle + \alpha\left\langle \mathrm{B}_\star\vec{1}, \mathrm{C}_\star\vec{1}\right\rangle\right|.$$

Proof By Proposition 2.14 (reducing to the linear term) we have

$$\Pr_{\sigma,\tau\sim S_n}[\sigma\in A, \tau\in B, \sigma\tau\in C] \ge 2\alpha\beta\gamma + \frac{2}{(n-1)^2}\langle \mathrm{BA}, \mathrm{C}\rangle - O\left(\frac{\sqrt{\alpha\beta\gamma}}{n}\right).$$

Our assumption implies $\alpha\beta\gamma \ge \delta^{-6}n^{-2}$, so the last term is $O(\delta^3\alpha\beta\gamma)$. Recall from Lemma 4.2 (reducing to the significant negative terms) that

$$\langle \mathrm{BA}, \mathrm{C}\rangle \ge \langle \mathrm{B}_-\mathrm{A}_{\mathrm{struc}}, \mathrm{C}_{\mathrm{struc}}\rangle + \langle \mathrm{B}_{\mathrm{struc}}\mathrm{A}_-, \mathrm{C}_{\mathrm{struc}}\rangle + \langle \mathrm{B}_{\mathrm{struc}}\mathrm{A}_{\mathrm{struc}}, \mathrm{C}_-\rangle - \sqrt{\delta}n^2(\log n)^{O(1)}\alpha\beta\gamma.$$

Applying Lemma 5.3 (lower bounding the negative entries) we obtain

$$\langle \mathrm{BA}, \mathrm{C}\rangle \ge -\beta\left\langle \mathrm{A}^t_{\mathrm{struc}}\vec{1}, \mathrm{C}^t_{\mathrm{struc}}\vec{1}\right\rangle - \gamma\left\langle \mathrm{A}_{\mathrm{struc}}\vec{1}, \mathrm{B}^t_{\mathrm{struc}}\vec{1}\right\rangle - \alpha\left\langle \mathrm{B}_{\mathrm{struc}}\vec{1}, \mathrm{C}_{\mathrm{struc}}\vec{1}\right\rangle$$

$$- \sqrt{\delta} n^2 (\log n)^{O(1)} \alpha\beta\gamma.$$

It remains to approximate the above matrix inner products by the corresponding terms involving star structure matrices. The calculations for the three terms are analogous, so we only show the details for the first term $-\beta \left\langle \mathrm{A}^t_{\mathrm{struc}} \overrightarrow{1}, \mathrm{C}^t_{\mathrm{struc}} \overrightarrow{1} \right\rangle$. We will apply Lemma 5.4 to $M = \mathrm{A}^t_{\mathrm{struc}}$ and $N = \mathrm{C}^t_{\mathrm{struc}}$. The level-1 inequality (Lemma 3.6) shows that its hypotheses are satisfied with $\eta_1 = \alpha n$ and $\eta_2 = \gamma n$. Thus we obtain

$$\begin{aligned} &\left| \left\langle \mathrm{A}^t_{\mathrm{struc}} \overrightarrow{1}, \mathrm{C}^t_{\mathrm{struc}} \overrightarrow{1} \right\rangle - \left\langle \mathrm{A}'_\star \overrightarrow{1}, \mathrm{C}'_\star \overrightarrow{1} \right\rangle \right| \\ &\quad \le \delta\alpha\gamma n^2 (\log n)^{O(1)} + \delta\alpha n \|\mathrm{C}'_\star\|_1 + \delta\gamma n \|\mathrm{A}'_\star\|_1. \end{aligned}$$

By Lemma 5.7 we have $\|\mathrm{A}'_\star\|_1 \le \alpha(1+2\delta)n$ and $\|\mathrm{C}'_\star\|_1 \le \gamma(1+2\delta)n$. Recalling that $\langle \mathrm{BA}, \mathrm{C} \rangle$ is multiplied by $\frac{2}{(n-1)^2}$ in the main calculation, we see that replacing $-\beta \left\langle \mathrm{A}^t_{\mathrm{struc}} \overrightarrow{1}, \mathrm{C}^t_{\mathrm{struc}} \overrightarrow{1} \right\rangle$ by $-\beta \left\langle \mathrm{A}'_\star \overrightarrow{1}, \mathrm{C}'_\star \overrightarrow{1} \right\rangle$ incurs an error term $\le \delta\alpha\beta\gamma (\log n)^{O(1)}$. Similar reasoning applies to the other terms, so the lemma follows. □

5.8 Product-free triples are somewhat explained by stars

We have now prepared the two main ingredients described in the proof overview earlier in the section: we have shown that the contributions from small associated stars are negligible and reduced the analysis of large associated stars to bounding the star structure terms that appear in Lemma 5.8. Our final lemma in preparation for the proof of Proposition 5.2, shows that the star structure terms are small except where they have a common row with a large sum, which corresponds to the associated stars with common centre required to prove Proposition 5.2.

Lemma 5.9 *Let $\eta_1, \eta_2 > 0$. Let $M_\star$, $N_\star$ be matrices with nonnegative coefficients and suppose that there is no coordinate i with both $(M_\star \overrightarrow{1})_i > \frac{\eta_1}{100}$ and $(N_\star \overrightarrow{1})_i > \frac{\eta_2}{100}$. Then*

$$\left\langle M_\star \overrightarrow{1}, N_\star \overrightarrow{1} \right\rangle \le \frac{\eta_2}{100} \|M_\star\|_1 + \frac{\eta_1}{100} \|N_\star\|_1.$$

Proof The terms of $\left\langle M_\star \overrightarrow{1}, N_\star \overrightarrow{1} \right\rangle$ that correspond to a coordinate in which $M_\star \overrightarrow{1}$ is $\le \frac{\eta_1}{100}$ sum up to at most $\frac{\eta_1}{100} \|N_\star \overrightarrow{1}\|_1$. In the rest of the terms the corresponding coordinate of $N_\star \overrightarrow{1}$ is at most $\eta_2/100$. We therefore have

$$\begin{aligned} \left\langle M_\star \overrightarrow{1}, N_\star \overrightarrow{1} \right\rangle &\le \frac{\eta_1}{100} \|N_\star \overrightarrow{1}\|_1 + \frac{\eta_2}{100} \|M^\star 1\|_1 \\ &= \frac{\eta_1}{100} \|N_\star\|_1 + \frac{\eta_2}{100} \|M^\star\|_1. \end{aligned}$$ □

We are now ready to prove Proposition 5.2.

Proof of Proposition 5.2 By Lemma 5.8 we have

$$\alpha\beta\gamma \le \frac{2}{n^2}\left(\beta\left\langle \mathrm{A}'_\star\vec{1}, \mathrm{C}'_\star\vec{1}\right\rangle + \gamma\left\langle \mathrm{A}_\star\vec{1}, \mathrm{B}'_\star\vec{1}\right\rangle + \alpha\left\langle \mathrm{B}_\star\vec{1}, \mathrm{C}_\star\vec{1}\right\rangle.\right).$$

If there is no coordinate i in which $(\mathrm{A}'_\star\vec{1})_i > \alpha n/100$ and $(\mathrm{C}'_\star\vec{1})_i > \gamma n/100$ then Lemma 5.9 and Lemma 5.7 bound $\left\langle \mathrm{A}'_\star\vec{1}, \mathrm{C}'_\star\vec{1}\right\rangle$ by $\|\mathrm{A}'_\star\|_1\gamma n/100 + \|\mathrm{C}'_\star\|_1\alpha n/100 < \frac{1}{40}\alpha\gamma n^2$. Similar calculations apply to the other two terms. However, these bounds cannot all hold, as then the inequality above would give $\alpha\beta\gamma < 3\cdot\frac{2}{40}\alpha\beta\gamma$, which is a contradiction. Thus we have the required coordinate i for one of the terms, e.g. for the third term this would give $(\mathrm{B}_\star\vec{1})_i > \beta n/100$ and $(\mathrm{C}_\star\vec{1})_i > \gamma n/100$, so $\mu_B(S_B(i)) > 1/100$ and $\mu_C(S_C(i)) > 1/100$, which corresponds to the stars described in the Proposition. □

5.9 A dense product-free set is explained by a single star

Now we will specialise the analysis of star structure terms from product-free triples to a single product-free set, where we will obtain the stronger structural conclusion that a single star explains the lack of products. We require the following lemma that bounds the star structure terms under the assumption that no single star accounts for almost all of the star structure matrices.

Lemma 5.10 *Let* $\zeta \in [0, 1/2]$ *and suppose that*

$$\max\left(\|\mathrm{A}_\star\vec{1}\|_\infty, \|\mathrm{A}'_\star\vec{1}\|_\infty\right) \le (1-\zeta)\left(\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1\right).$$

Then

$$\left\langle \mathrm{A}_\star\vec{1}, \mathrm{A}_\star\vec{1}\right\rangle + \left\langle \mathrm{A}'_\star\vec{1}, \mathrm{A}'_\star\vec{1}\right\rangle + \left\langle \mathrm{A}'_\star\vec{1}, \mathrm{A}_\star\vec{1}\right\rangle \le (1-\zeta/2)\left(\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1\right)^2.$$

Proof The lemma is immediate from the following claim applied to $v = \frac{\mathrm{A}_\star\vec{1}}{\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1}$ and $u = \frac{\mathrm{A}'_\star\vec{1}}{\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1}$.

Claim 5.11 *Let* $v, u \in [0, 1-\zeta]^n$ *with* $\|v\|_1 + \|u\|_1 = 1$. *Then*

$$\|v\|_2^2 + \|u\|_2^2 + \langle v, u\rangle \le 1 - \zeta(1-\zeta).$$

To see this, suppose that $v_1 + u_1 \ge v_i + u_i$ for all other i. Then we have

$$\begin{aligned}\|v\|_2^2 + \|u\|_2^2 + \langle v, u\rangle &= \sum_{i=1}^{n}\left((v_i+u_i)^2 - v_iu_i\right)\\ &\le \sum_{i=1}^{n}(v_1+u_1)(v_i+u_i) - v_1u_1\end{aligned}$$

$$= v_1 + u_1 - v_1 u_1.$$

The function $f(u_1, v_1) := v_1 + u_1 - v_1 u_1$ is increasing in both coordinates, so is maximised when $u_1 + v_1 = 1$. Now $g(v_1) := f(1 - v_1, v_1) = 1 - v_1(1 - v_1)$ is a convex function of v_1, so is maximised at the boundary of its domain $[\zeta, 1 - \zeta]$. This proves the claim, and the lemma follows. □

Now we can show for any moderately dense product-free set $A \subseteq A_n$ that some associated star or inverse star explains A up to a factor $1 + O(\delta^{1/4})$.

Lemma 5.12 *Suppose that $A \subseteq A_n$ is product-free with $\mu(A) \geq \delta^{-2} n^{-2/3}$. Then $\max_{i \in [n]} \max\{s_A(i), s'_A(i)\} = (1 + O(\delta^{1/4}))\alpha$.*

Proof We define $\zeta' \in [0, 1]$ by

$$\max\left(\|\mathrm{A}_\star \vec{1}\|_\infty, \|\mathrm{A}'_\star \vec{1}\|_\infty\right) = (1 - \zeta')(\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1)$$

and let $\zeta = \min(\zeta', 1/2)$. Applying Lemma 5.8 with $A = B = C$, then Lemma 5.10, and then Lemma 5.7, we obtain

$$\begin{aligned}(n-1)^2 \alpha^3 \left(1 - O\left(\delta^{1/4}\right)\right) &\leq \alpha \left(\left\langle \mathrm{A}'_\star \vec{1}, \mathrm{A}'_\star \vec{1} \right\rangle + \left\langle \mathrm{A}_\star \vec{1}, \mathrm{A}'_\star \vec{1} \right\rangle + \left\langle \mathrm{A}_\star \vec{1}, \mathrm{A}_\star \vec{1} \right\rangle\right) \\ &\leq \alpha \left(\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1\right)^2 (1 - \zeta/2) \\ &\leq n^2 \alpha^3 (1 - \zeta/2)(1 + O(\delta)).\end{aligned}$$

Thus $\zeta = O(\delta^{1/4})$, so $\zeta' = \zeta$, and $\|\mathrm{A}_\star\|_1 + \|\mathrm{A}'_\star\|_1 = (1 + O(\delta^{1/4}))\alpha n$, so by definition of ζ' we have

$$(n-1) \max_{i \in [n]} \max\{s_A(i), s'_A(i)\} = \max\left(\|\mathrm{A}'_\star \vec{1}\|_\infty, \|\mathrm{A}'_\star \vec{1}\|\right) = (1 + O(\delta^{1/4}))\alpha n.$$

This completes the proof. □

5.10 A dense product-free set is close to a single star

Now we will refine the star structure obtained in the previous subsection to prove Proposition 5.1, which shows that any moderately dense product-free set is closely approximated by a single star. Our idea is to use the (inverse) star S provided above with the fact that $(A, A, A \setminus S)$ is a product-free triple to which we can apply Proposition 5.2 to deduce that $A \setminus S$ is small.

Proof of Proposition 5.1 By Lemma 5.12 there is a large associated (inverse) star S_1 such that $\mu(A \setminus S_1) < O(\delta^{1/4})\mu(A)$. Without loss of generality $S_1 = 1_{1 \to I_1}$. Let

$$I = \left\{i : \mu(A_{1 \to i}) \geq n^{-\frac{1}{3}}\right\},$$

let $S = 1_{1 \to I}$, and let $C = A \setminus S$.

We assert that $\mu(C) \leq \delta^{-2} n^{-2/3}$. Suppose otherwise. Then the triple (A, A, C) is product-free, and we can apply Proposition 5.2 to deduce that its conclusion holds for some triple equivalent to (A, A, C). In principle, the possibilities are for some i that

1. $\mu_A(S_A(i)) > 1/100$ and $\mu_A(S'_A(i)) > 1/100$, or
2. $\mu_A(S'_A(i)) > 1/100$ and $\mu_C(S'_C(i)) > 1/100$, or
3. $\mu_A(S_A(i)) > 1/100$ and $\mu_C(S_C(i)) > 1/100$.

However, $\mu_A(S_A(1))$ is so large that it precludes the existence of any other associated star or inverse star S with $\mu_A(S) > 1/100$, so the only possibility is that (3) holds with $i = 1$. However, by definition of C we have $\mu\left(C_{1\to j}\right) \leq n^{-\frac{1}{3}}$ for all i, but as $S_C(1) \neq \varnothing$ we can choose j with

$$\mu(C_{1\to j}) > c_{1j} > \epsilon_C = n\delta\mu(C)\mu(A) > n^{-\frac{1}{3}}.$$

This contradiction completes the proof. □

5.11 Product-free triples when one set has no large associated stars

We have now completed the proofs for the main goals of the section. For future reference we will conclude the section by proving a star-structure theorem for product-free triples (A, B, C) where A has no large associated stars and inverse stars. The rationale is that after reordering we expect B and C to look like stars and A to look like $1_{I\to\overline{J}}$. We therefore expect A to have no large associated stars. Under this assumption, we strengthen the conclusion of Proposition 5.2 (product-free triples are somewhat explained by stars) to the same level of accuracy that we achieved for a single product-free set: we show that B and C are each explained up to a factor $1 + O(\delta^{1/4})$ by a single star with the same centre.

Lemma 5.13 *Suppose that (A, B, C) are as in Proposition 5.2 and that A has no associated stars and inverse stars. Then there exists $i \in [n]$ such that $\mu(B \setminus S_B(i)) \leq O(\delta^{1/4})\beta$ and $\mu(C \setminus S_C(i)) \leq O(\delta^{1/4})\gamma$.*

Proof The proof is similar to that of Lemma 5.12. By assumption we have $\mathrm{A}_\star$ and $\mathrm{A}'_\star = 0$, so Lemma 5.8 reduces to

$$(n-1)^2\alpha\beta\gamma\left(1 - O\left(\delta^{1/4}\right)\right) = \alpha\left\langle \mathrm{B}_\star\vec{1}, \mathrm{C}_\star\vec{1}\right\rangle.$$

Let $v = \frac{\mathrm{B}_\star\vec{1}}{\|\mathrm{B}_\star\|_1}$ and $u = \frac{\mathrm{C}_\star\vec{1}}{\|\mathrm{C}_\star\|_1}$. Then $\langle v, u\rangle > 1 - O(\delta^{1/4})$, as otherwise Lemma 5.7 would give

$$(n-1)^2\alpha\beta\gamma < (1-\Omega(\delta^{1/4}))\alpha\|\mathrm{B}_\star\|_1\|\mathrm{C}_\star\|_1 < (1-\Omega(\delta^{1/4}))\alpha(1+2\delta)\beta n(1+2\delta)\gamma n,$$

which is a contradiction. Now in the place of Claim 5.11 we must show that there is some coordinate i with both v_i and u_i at least $1 - O\left(\delta^{1/4}\right)$. As

$$\langle v, u\rangle \leq \|v\|_\infty\|u\|_1 = \|v\|_\infty$$

we have $\|v\|_\infty \geq 1 - O\left(\delta^{1/4}\right)$ and similarly $\|u\|_\infty \geq 1 - O\left(\delta^{1/4}\right)$, so each of u and v has one coordinate equal to $1 - O\left(\delta^{1/4}\right)$ with the others all $O\left(\delta^{1/4}\right)$. It remains to note that $\|v\|_\infty$ and $\|u\|_\infty$ must be achieved at the same coordinate, as otherwise we would have

$$\langle v, u\rangle \leq O\left(\delta^{1/4}\right)\left(\|v\|_1 + \|u\|_1\right) = O\left(\delta^{1/4}\right),$$

which contradicts $\langle v, u\rangle > 1 - O(\delta^{1/4})$. □

6 Bootstrapping

In this section we will use the star structure established in the previous section to prove our main results, which give exact extremal results and strong stability results for product-free sets in A_n. The proofs will use the large restrictions provided by the star structure to deduce that other restrictions must be much smaller. This will allow us to progressively tighten our approximate structure until it becomes exact.

6.1 Product-free restrictions

As discussed in Sect. 1.5, if some product-free $A \subseteq A_n$ is well approximated by a star $1_{x\to I}$ then for each $i, i' \in I$ we will see that A has small density in $\mathcal{D}_{i\to i'}$ by inspecting the triple $(A_{i\to i'}, A_{x\to i}, A_{x\to i'})$ and factoring out the corresponding dictators. This is formalised by the following lemma.

Lemma 6.1 *Let $\epsilon > \frac{1}{\delta\sqrt{n}}$, and let (A, B, C) be product-free. Suppose that $\mu\left(B_{i\to j}\right) \geq \epsilon$ and $\mu\left(C_{i\to k}\right) \geq \epsilon$. Then*

$$\mu\left(A_{j\to k}\right) \leq \frac{\log^{O(1)}(1/\epsilon)}{\epsilon n}.$$

Proof We apply the following transformation that preserves products: let $B' = (jn)\,B_{i\to j}\,(ni)$, $A' = (nk)\,A_{j\to k}\,(nj)$ and $C' = (nk)\,C_{i\to k}\,(ni)$. Then the equation $ab = c$ in $\left(A', B', C'\right)$ is equivalent to the corresponding equation inside $\left(A_{j\to k}, B_{i\to j}, C_{i\to k}\right)$. As $\left(A', B', C'\right)$ can be viewed as product-free subsets of A_{n-1} the lemma follows from Corollary 3.10. □

The following lemma shows that if (A, B, C) are product-free and B, C are dense in stars $1_{x\to I}$, $1_{x\to J}$ then A is sparse outside of $1_{I\to\overline{J}} := \{\sigma : \sigma(I) \cap J = \varnothing\}$.

Lemma 6.2 *Let $\epsilon > \frac{1}{\delta\sqrt{n}}$ and let (A, B, C) be a product-free triple. Let $x \in [n]$ and let $I, J \subseteq [n]$. Suppose that for each $i \in I$ and each $j \in J$ we have $\mu\left(B_{x\to i}\right), \mu\left(C_{x\to j}\right) \geq \epsilon$. Suppose further that $\mu(A) \geq \frac{1}{\delta\epsilon n}$. Then $|I|\,|J| \leq 10n\log n$ and*

$$\mu\left(A \setminus 1_{I\to\overline{J}}\right) \leq \frac{|I|\,|J|\log^{O(1)}(1/\epsilon)}{\epsilon n^2}.$$

Proof By a union bound and Lemma 6.1 we have

$$\mu\left(A \setminus 1_{I \to \overline{J}}\right) \leq \sum_{i \in I, j \in J} \frac{1}{n} \mu\left(A_{i \to j}\right) \leq \frac{|I||J|}{n} \frac{\log^{O(1)}(1/\epsilon)}{\epsilon n}. \tag{6}$$

To complete the proof we show that $|I||J| \leq 10n \log n$. Suppose otherwise. Then by removing elements from either I or J we may assume that $|I||J| \in (10n \log n, 20n \log n)$. In which case we have

$$\mu\left(1_{I \to \overline{J}}\right) = \left(1 - \frac{|J|}{n}\right)\left(1 - \frac{|J|}{n-1}\right) \cdots \left(1 - \frac{|J|}{n - |I|}\right)$$
$$\leq \left(1 - \frac{|J|}{n}\right)^{|I|} \leq e^{-|J||I|/n} \leq n^{-3},$$

Together with (6) this yields

$$\mu(A) \leq \frac{\log^{O(1)} n}{\epsilon n} + n^{-3},$$

which contradicts the hypothesis $\mu(A) \geq \frac{1}{\delta \epsilon n}$. This shows that $|I||J| \leq 10n \log n$. □

6.2 Stability result for product-free sets

We now prove the following stronger version of Theorem 1.2.

Theorem 6.3 *Suppose that A is product-free with $\mu(A) \geq \delta^{-2} n^{-2/3}$. Then up to inversion there exist $x \in [n]$ and $I \subseteq [n]$ with $|I|^2 \leq 10n \log n$ such that $\mu(A_{x \to i}) \geq n^{-\frac{1}{3}}$ for each $i \in I$ and*

$$\mu\left(A \setminus F_I^x\right) \leq O\left(\delta^{-2}\right) n^{-2/3}.$$

Proof By Proposition 5.1, up to inversion there exist $x \in [n]$ and $I \subseteq [n]$ such that for each $i \in I$ we have $\mu(A_{x \to i}) \geq n^{-1/3}$ and $\mu(A \setminus 1_{x \to I}) \leq O(\delta^{-2}) n^{-\frac{2}{3}}$. Without loss of generality we assume this for A rather than A^{-1}. By Lemma 6.2 with $A = B = C$ and $\epsilon = n^{-\frac{1}{3}}$ we have $|I|^2 \leq 10n \log n$ and

$$\mu\left(A \setminus 1_{I \to \overline{I}}\right) \leq \frac{\log^{O(1)} n}{n^{2/3}}.$$

This shows that

$$\mu\left(A \setminus F_I^x\right) \leq \mu(A \setminus 1_{x \to I}) + \mu\left(A \setminus 1_{I \to \overline{I}}\right) \leq O\left(\delta^{-2}\right) n^{-\frac{2}{3}}.$$ □

6.3 Bootstrapping triples

With slightly more work we are also able to prove the following stability result for product-free triples that are somewhat sparse (we replace a log factor by a log log factor).

Theorem 6.4 *There is an absolute constant K such that if n is sufficiently large and (A, B, C) is a product-free triple in A_n with*

$$\min(\alpha\beta, \beta\gamma, \gamma\alpha) \geq \frac{(\log\log n)^K}{n}$$

then up to equivalence there exist $I, J \subseteq [n], x \in [n]$ with $|I||J| \leq 10n\log n$ such that

$$\mu_B(1_{x\to I}), \mu_C(1_{x\to J}) \geq 1 - O\left(\delta^{1/4}\right) \quad \text{and} \quad \mu\left(A \setminus 1_{I\to\overline{J}}\right) \leq \delta^{-2}n.$$

In particular, we have

$$\mu\left(A \setminus 1_{I\to\overline{J}}\right) = o(\alpha), \quad \mu(B\setminus 1_{x\to I}) = o(\beta), \quad \mu(C\setminus 1_{x\to J}) = o(\gamma).$$

The idea of the proof is to start with the weaker star structure guaranteed from Proposition 5.2. Lemma 6.2 will then easily imply a weaker variant of the theorem with $1 - O\left(\delta^{1/4}\right)$ replaced by $\frac{1}{100}$. This will allow us to show that actually A has no large associated stars and inverse stars, so instead of Proposition 5.2 we may apply the more suitable Lemma 5.13.

Proof We claim that $\alpha \geq \frac{1}{\delta^{100}n}$. Indeed, otherwise our assumption implies $\epsilon := \min(\beta, \gamma) > (\log\log n)^K \delta^{100}$, and then Corollary 3.10 gives $\alpha \leq \frac{\log^{O(1)}(1/\epsilon)}{\epsilon n}$, so recalling $\delta^{-1} = \log^R n$ we have $\min(\alpha\beta, \alpha\gamma) < \frac{(\log\log n)^{O(1)}}{n}$, which contradicts our assumption for K large enough. Thus we have the claimed lower bound on α.

Similarly, we have the same bound on β and γ. Combining these with $\min(\alpha\beta, \beta\gamma, \gamma\alpha) \geq \frac{(\log\log n)^K}{n}$ we deduce the assumption of Proposition 5.2. By this Proposition, up to equivalence of the triple (A, B, C) there exist associated stars $S_B(x) = 1_{x\to I}$, $S_C(x) = 1_{x\to J}$, such that $\mu_B(S_B(x)) > \frac{1}{100}$ and $\mu_C(S_C(x)) \geq \frac{1}{100}$. Our assumptions imply $\min(\epsilon_B, \epsilon_C) \geq \delta$, so by Lemma 6.2 we have $|I||J| \leq 10n\log n$ and

$$\mu\left(A \setminus 1_{I\to\overline{J}}\right) \leq \frac{1}{\delta^2 n}$$

We now assert that A has no large associated star or inverse star S. Indeed, for such S, as $\mu(A \cap S) \geq \delta\mu(A)$ we would have

$$\mu(S) \leq \frac{\mu(A)}{\epsilon_A} \leq \frac{\mu(A)}{\delta} \leq \frac{\mu(A \cap S)}{\delta^2}.$$

On the other hand,

$$\mu(A \cap S) \leq \mu\left(S \cap 1_{I\to\overline{J}}\right) + \mu\left(A \setminus 1_{I\to\overline{J}}\right)$$

$$\leq 2\mu(S)\, e^{-\frac{|I||J|}{n}} + \frac{1}{\delta^2 n}.$$

We have

$$\frac{|I||J|}{n^2} \geq \mu(S_B(x))\mu(S_C(x)) \geq \frac{\beta}{100}\frac{\gamma}{100} > \frac{(\log\log n)^K}{10^4 n},$$

so

$$\mu(A) \leq \delta^{-1}\mu(A \cap S) \leq 2\mu(S)\,\delta^8 + \frac{1}{\delta^3 n} \leq 2\mu(A)\,\delta^6 + \frac{1}{\delta^3 n}.$$

This contradicts the bound $\mu(A) = \alpha \geq \frac{1}{\delta^{100} n}$, so such S cannot exist.

Thus we can apply Lemma 5.13 to obtain the required stronger approximation of B and C by associated stars, i.e. $\mu_B(S_B(x))$ and $\mu_C(S_C(x))$ are both $1 - O\left(\delta^{1/4}\right)$. □

6.4 Further bootstrapping of product-free sets

Recall from Theorem 6.3 that if A is a dense product-free set then $\mu\left(A \setminus F_I^x\right)$ is small for some x and I. If A is extremal this implies $\mu_{F_I^x}(A) \approx 1$. Our goal in this subsection is to show for such A that

$$\mu\left(A \setminus F_I^x\right) \leq \frac{1}{2}\mu\left(F_I^x \setminus A\right)$$

and therefore any extremal product-free A must be of the form F_I^x. Our proof will also work for sets that are sufficiently close to extremal.

We require various lemmas that will be applied to certain restrictions of A. The first considers a product-free triple (A, B, C) (which will be restrictions of the original A) and shows that if two sets are dense in sets of the form $1_{I \to \overline{I}}$ then the third must be empty.

Lemma 6.5 *Let $I_1, I_2, J_1, J_2 \subseteq [n]$ with n sufficiently large and $|I_1| + |I_2| + |J_1| + |J_2| \leq 40\sqrt{n}$. Suppose that (A, B, C) is product-free. If $\mu_{I_1 \to \overline{J_1}}(C)$ and $\mu_{I_2 \to \overline{J_2}}(B)$ are both at least $1 - e^{-2000}$ then A is empty. Similar statements hold for all permutations of ABC.*

Proof We only prove the first statement, as the proofs for permutations of ABC are similar. Suppose on the contrary that there exists $\tau \in A$. Let $\sigma \sim S_n$ be a random permutation. We will derive a contraction by showing that $\Pr[\sigma \in B, \tau\sigma \in C] > 0$. By assumption, we have

$$\Pr[\sigma \in B, \tau\sigma \in C] \geq \Pr\left[\sigma(I_1) \subseteq \overline{J_1}, \tau\sigma(I_2) \subseteq \overline{J_2}\right] - 2e^{-2000}.$$

On the other hand,

$$\Pr\left[\sigma(I_1) \subseteq \overline{J_1}, \tau\sigma(I_2) \subseteq \overline{J_2}\right] = \mu\left(1_{I_1 \to \overline{J_1}} \cap 1_{I_2 \to \tau^{-1}(\overline{J_2})}\right)$$

$$\geq \left(1 - \frac{|I_1| + |I_2| + |J_1| + |J_2|}{n}\right)^{|I_1|+|I_2|}$$
$$> 2e^{-2000},$$

provided that n is sufficiently large. Indeed, the second inequality follows from defining a random σ on the indices $i \in I_1 \cup I_2$ one by one, noting that in each step we have at least $n - |I_1| - |I_2| - |J_1| - |J_2|$ free options. Thus $\Pr[\sigma \in B, \tau\sigma \in C] > 0$, contrary to our assumption that (A, B, C) is product-free, so $A = \varnothing$. □

We now show that if (A, B, C) is a product-free triple, B contains almost all of $1_{I_1 \to \overline{J_1}}$ and A contains almost all of F_I^x then C is small.

Lemma 6.6 *Let $\zeta \in \left(0, e^{-2000}\right)$ and let (A, B, C) be a product-free triple in A_n with n sufficiently large. Suppose for some I_1, J_1, I of size $\leq 10\sqrt{n}$ that $\mu_{I_1 \to \overline{J_1}}(B) \geq 1 - \zeta$ and $\mu_{F_I^x}(A) \geq 1 - \frac{1}{2}e^{-2000}$. Then*

$$\mu(C) \leq \left(e^{2000}\zeta\right)^{\frac{|I|}{2}}.$$

Moreover, if $\zeta < \frac{|I|}{2e^{2000}n}$ then C is empty.

Proof Let

$$I_2 := \{i \in I : \mu_{I \to \overline{I}}(A_{x \to i}) \geq 1 - e^{-2000}\}.$$

By definition of F_I^x, our assumption on A and Markov's inequality we have $|I_2| \geq \frac{|I|}{2}$. Similarly, letting

$$J := \{j \in [n] : \mu_{I_1 \to \overline{J_1}}\left(B_{j \to x}\right) \geq 1 - e^{-2000}\},$$

Markov's inequality applied to B gives $|J| \geq \left(1 - e^{2000}\zeta\right)n$.

For each $i \in I_2$ and $j \in J$ we can apply Lemma 6.5 to the triple $(A_{x \to i}, B_{j \to x}, C_{j \to i})$ to conclude that each such $C_{j \to i}$ is empty, so $C \subseteq 1_{J \to \overline{I_2}}$. This shows that

$$\mu(C) \leq \mu\left(1_{J \to \overline{I_2}}\right) = \prod_{i=1}^{|I_2|} \frac{\left|\overline{J}\right| - i}{n - i} \leq \left(1 - \frac{\left|\overline{J}\right|}{n}\right)^{|I_2|} \leq \left(e^{2000}\zeta\right)^{|I|/2}.$$

Finally, if $\zeta < \frac{|I|}{2e^{2000}n}$ then $\left|\overline{I_2}\right| < |J|$, and so $C \subseteq 1_{J \to \overline{I_2}} = \varnothing$. □

After the above preliminary lemmas, we now come to the main engine of our bootstrapping approach, which shows that if $\mu\left(A \setminus F_I^x\right)$ is somewhat small then it is much smaller than $\mu\left(F_I^x \setminus A\right)$.

Lemma 6.7 *Let $A \subseteq A_n$ be product-free with n sufficiently large. Suppose that*

$$I := \{i \in [n] : \mu(A_{1 \to i}) \geq n^{-\frac{1}{3}}\}.$$

satisfies $\frac{\sqrt{n}}{10} < |I| \leq 10\sqrt{n}$. *Suppose* $\mu_{F_I^x}(A) \geq 1 - \zeta$ *with* $\zeta \in \left(0, e^{-20000}\right)$. *Then*

$$\mu\left(A \setminus F_I^x\right) \leq \zeta n^{-2/3} \log^{O(1)} n.$$

Moreover, if $\zeta \leq \frac{1}{e^{8000}\sqrt{n}}$ *then* $A \subseteq F_I^x$.

Proof We define $I_2 \subseteq I_1 \subseteq I$ by

$$I_1 = \{i \in I : \mu_{I \to \overline{I}}(A_{x \to i}) \geq 1 - e^{-2000}\} \quad \text{and}$$
$$I_2 = \{i \in I_1 : \mu_{I \to \overline{I}}(A_{x \to i}) \geq 1 - 2\zeta\}.$$

By Markov's inequality we have $|I \setminus I_1| \leq e^{2000}\zeta\, |I|$ and $|I_2| \geq \frac{1}{2}|I|$.

Let $J \subseteq \overline{I}$ consist of the $\lfloor 80 \cdot e^{4000}\zeta\sqrt{n} \rfloor$ indices $i \in \overline{I}$ in which $\mu(A_{x\to i})$ is largest. We partition $\left(A \setminus F_I^x\right)$ into the following three bad events and bound each of their measures separately.

1. Let $B_1 = \{\sigma \in A : \sigma(x) \in I \text{ and } \sigma(I) \cap I \neq \varnothing\}$.
2. Let $B_2 = \{\sigma \in A : \sigma(x) \in J\}$.
3. Let $B_3 = \{\sigma \in A : \sigma(x) \in \overline{I \cup J}\}$.

To prove the lemma, it suffices to show for each $i \in [3]$ that $\mu(B_i) < \zeta n^{-2/3} \log^{O(1)} n$, and if moreover $\zeta \leq \frac{1}{e^{8000}\sqrt{n}}$ then $B_i = \varnothing$.

Upper bounding $\boldsymbol{\mu(B_1)}$. By a union bound we have

$$\mu(B_1) \leq \sum_{i,j \in I} \frac{\mu\left(A_{i\to j}\right)}{n}.$$

By definition of I_1, for each $i, j \in I_1$ we can apply Lemma 6.5 to $(A_{x\to j}, A_{i\to j}, A_{x\to i})$ to obtain $A_{i\to j} = \varnothing$. For general $i, j \in I$ we may apply Lemma 6.1 to obtain

$$\mu\left(A_{i\to j}\right) \leq n^{-\frac{2}{3}} \log^{O(1)} n.$$

Therefore,

$$\mu(B_1) \leq \sum_{i,j \in I} \frac{\mu\left(A_{i\to j}\right)}{n} \leq 2|I \setminus I_1|\,|I|\, n^{-5/3} \log^{O(1)} n \leq \zeta n^{-2/3} \log^{O(1)} n.$$

Moreover, if $\zeta \leq \frac{1}{e^{8000}\sqrt{n}}$ then $|I \setminus I_1| \leq e^{2000}\zeta|I| < e^{-6000}|I|/\sqrt{n} < 1$, i.e. $I_1 = I$ and so $B_1 = \varnothing$.

Upper bounding $\boldsymbol{\mu(B_2)}$. By definition of I, a simple union bound gives

$$\mu(B_2) \leq \sum_{j \in J} \frac{\mu\left(A_{x\to j}\right)}{n} \leq \frac{|J|}{n} n^{-1/3} = O(\zeta n^{-5/6}).$$

Moreover, if $\zeta \leq \frac{1}{e^{8000}\sqrt{n}}$ then $|J| < 1$, i.e. $J = \varnothing$ and so $B_2 = \varnothing$,

Upper bounding $\mu(B_3)$. Here we will use the bound $\mu(B_3) \le \mu\left(A_{x\to j}\right)$, where we fix $j \in \overline{I \cup J}$ to maximise $\mu\left(A_{x\to j}\right)$.

It suffices to show that $\mu\left(A_{x\to j}\right) \le \left(2e^{2000}\zeta\right)^{\frac{\sqrt{n}}{20}}$, and moreover that if $\zeta \le \frac{1}{e^{8000}\sqrt{n}}$ then $A_{x\to j} = \varnothing$. We suppose for a contradiction that this fails.

We will consider $J' := J \cup \{j\}$, noting that $|J'| = \lfloor 80 \cdot e^{4000}\zeta\sqrt{n}\rfloor + 1$ and for each $j' \in J'$ we have $\mu\left(A_{x\to j'}\right) \ge \mu\left(A_{x\to j}\right)$.

We will apply Lemma 6.6 to the product-free triple $\left(A_{i\to j'}, A_{x\to i}, A_{x\to j'}\right)$ for each $i \in I_2$ and $j' \in J'$. By our above assumption on $C := A_{x\to j}$ the conclusion of this lemma does not hold, so one of the hypotheses does not hold. By definition of I_2, the hypothesis for $B := A_{x\to i}$ is satisfied with 2ζ in place of ζ. Thus the hypothesis for $A_{i\to j'}$ does not hold, so

$$\mu_{F_I^x}\left(A_{i\to j'}\right) \le 1 - \tfrac{1}{2}e^{-2000}.$$

Inclusion–Exclusion then shows that

$$\begin{aligned}\mu_{F_I^x}\left(\overline{A}\right) &\ge \sum_{i\in I_2, j'\in J'} e^{-4000}\mu_{F_I^x}\left(1_{i\to j'}\right) - \sum_{\substack{i_1,i_2\in I_2\\ j_1,j_2\in J'}} \mu_{F_I^x}\left(1_{i_1\to j_1} \cap 1_{i_2\to j_2}\right)\\ &\ge e^{-4000}|I_2||J'|/n - 2|I_2|^2|J'|^2/n^2 \ge e^{-4000}|I_2||J'|/2n.\end{aligned}$$

However, by assumption $\mu_{F_I^x}\left(\overline{A}\right) \le \zeta$, so

$$|J'| \le 2e^{4000}\zeta n/|I_2| \le 4e^{4000}\zeta n/|I| < 40e^{4000}\zeta\sqrt{n},$$

which contradicts the choice of $|J'|$. Thus the required bound for $\mu\left(A_{x\to j}\right)$ holds, so the lemma follows. □

We conclude this section with the proof of Theorem 1.3, which implies Theorem 1.1 (our main theorem).

Proof of Theorem 1.3 We introduce the absolute constant $c = e^{-9000}$. We need to show that if n is sufficiently large and $A \subseteq A_n$ is a product-free set with $\mu(A) > \max_{I,x}\mu\left(F_I^x\right) - \frac{c}{n}$ then up to inversion A is contained in some F_I^x. By Theorem 6.3 there exists some F_I^x such that $\mu\left(A \setminus F_I^x\right) \le n^{-0.66}$, where we may assume that this holds for A rather than A^{-1}, and moreover $\mu\left(A_{x\to i}\right) \ge n^{-\frac{1}{3}}$ for each $i \in I$. We note that $\mu(F_I^x) \sim \mu(A) \sim 1/\sqrt{2en}$, so $\frac{\sqrt{n}}{10} < |I| < 10\sqrt{n}$. Writing $\mu_{F_I^x}(A) = 1 - \zeta$ we have $\zeta = O(n^{-0.16}) < e^{-20000}$.

Thus we can apply Lemma 6.7 to obtain

$$\mu\left(A \setminus F_I^x\right) \le \zeta n^{-2/3}\log^{O(1)} n < \zeta\mu(F_I^x)/2.$$

We deduce

$$\mu\left(F_I^x\right) - \frac{c}{n} \le \mu(A) \le \mu\left(F_I^x\right)\mu_{F_I^x}(A) + \mu\left(A \setminus F_I^x\right) \le \mu\left(F_I^x\right)(1 - \zeta/2).$$

By choice of c this implies $\zeta \leq \frac{1}{e^{8000}\sqrt{n}}$, so we can apply Lemma 6.7 to obtain $A \subseteq F_I^x$. □

7 Concluding remarks

Our methods for bounding product-free subsets of the alternating group have justified the intuitive idea that globalness can be regarded as a pseudorandomness notion, by showing estimates on the eigenvalues of Cayley operators for global sets that correspond to the intuition for random sets. Given the large literature on random Cayley graphs inspired by the seminal paper of Alon and Roichman [1], one natural direction for further research is whether analogous results hold for Cayley graphs with respect to global sets.

We also propose the study of extremal problems for word maps in general groups (see the survey of Shalev [16] for background). Fix any group G. Any word $w = w(x_1, \ldots, x_d)$ in the free group F_d on d generators naturally defines a word map $w : G^d \to G$. For example, if $d = 3$, $w = x_1 x_2 x_3^{-1}$, and $A \subseteq G$ then A is product-free iff $A^3 \cap \ker w = \varnothing$. Our main result therefore describes the largest $A \subseteq A_n$ such that $A^3 \cap \ker w = \varnothing$, and so suggests the following more general problem.

Problem 7.1 For any finite group G and word $w \in F_d$, what is the largest $A \subseteq G$ with $A^d \cap \ker w = \varnothing$?

Acknowledgements We thank an anonymous referee, Anubhab Ghosal and Ferdinand Jacobé de Naurois for reading the paper carefully.

Funding Open Access funding provided by the MIT Libraries. P. K. Supported by ERC Advanced Grant 883810. D. M. Supported by a Sloan Research Fellowship.

References

1. Alon, N., Roichman, Y.: Random Cayley graphs and expanders. Random Struct. Algorithms **5**, 271–284 (1994)
2. Babai, L., Sós, V.T.: Sidon sets in groups and induced subgraphs of Cayley graphs. Eur. J. Comb. **6**(2), 101–114 (1985)
3. Bourgain, J., Gamburd, A.: Expansion and random walks in $SL_d(\mathbb{Z}/p\mathbb{Z})$: I. J. Eur. Math. Soc. **10**(4), 987–1011 (2008)
4. Eberhard, S.: Product mixing in the alternating group. Discrete Anal. **2**, 18 (2016)
5. Ellis, D., Friedgut, E., Pilpel, H.: Intersecting families of permutations. J. Am. Math. Soc. **24**(3), 649–682 (2011)

6. Filmus, Y., Kindler, G., Lifshitz, N., Minzer, D.: Hypercontractivity on the symmetric group (2020). ArXiv preprint. arXiv:2009.05503
7. Friedgut, E.: On the measure of intersecting families, uniqueness and stability. Combinatorica **28**(5), 503–528 (2008)
8. Gowers, W.T.: Quasirandom groups. Comb. Probab. Comput. **17**(3), 363–387 (2008)
9. Green, B.: 100 open problems. Manuscript
10. Keevash, P., Lifshitz, N., Long, E., Minzer, D.: Global hypercontractivity and its applications (2021). ArXiv preprint. arXiv:2103.04604
11. Kiran, K.S.: Large product-free subsets of finite groups. J. Comb. Theory, Ser. A **77**(2), 339–343 (1997)
12. Nikolov, N., Pyber, L.: Product decompositions of quasirandom groups and a Jordan type theorem. J. Eur. Math. Soc. **13**(4), 1063–1077 (2011)
13. O'Donnell, R.: Analysis of Boolean Functions. Cambridge University Press, Cambridge (2014)
14. Persi, D.: Group Representations in Probability and Statistics. Lecture Notes-Monograph Series, pp. 1–192 (1988)
15. Sarnak, P., Xue, X.: Bounds for multiplicities of automorphic representations. Duke Math. J. **64**(1), 207–227 (1991)
16. Shalev, A.: Some results and problems in the theory of word maps. Erdős Centen.: Bolyai Soc. Math. Stud., 611–649 (2013)